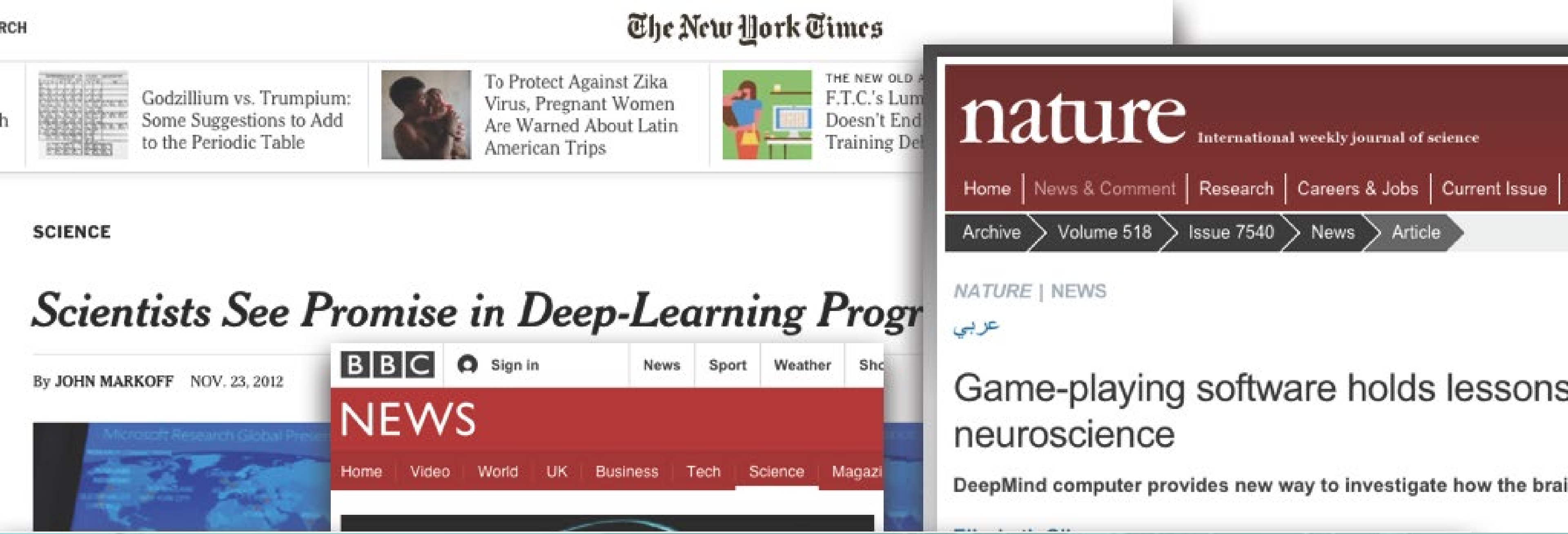
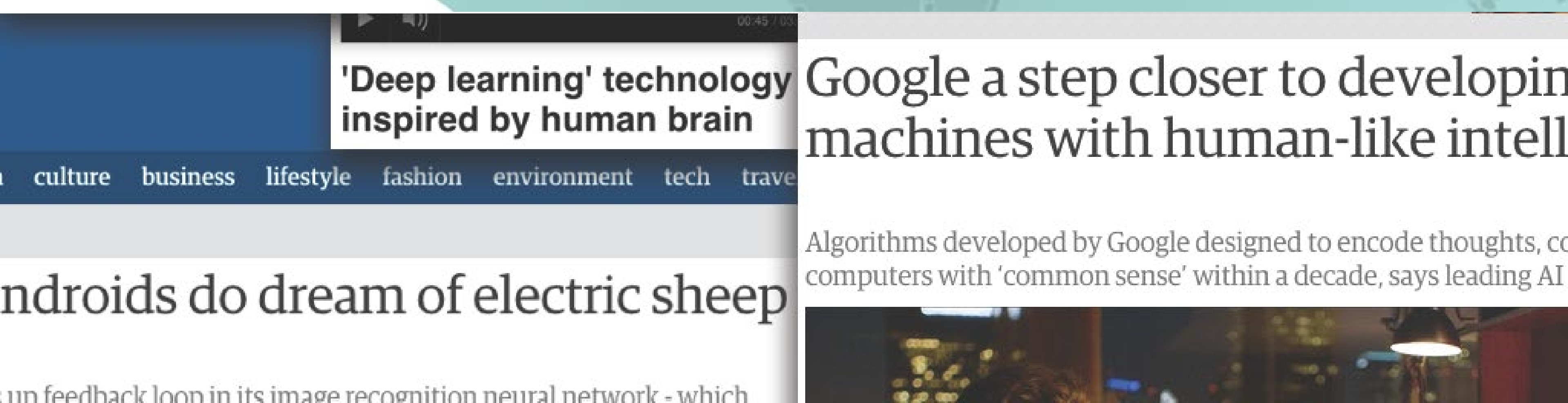
(L,W)CS182/282A: Designing, Visualizing and Understanding Deep Neural Networks

John Canny

Spring 2020



Did Facebook Shutdown An AI That Made Its Own Language? AI Will Never Replace Humans and Artificial Intelligence's Threat may Already Be Here



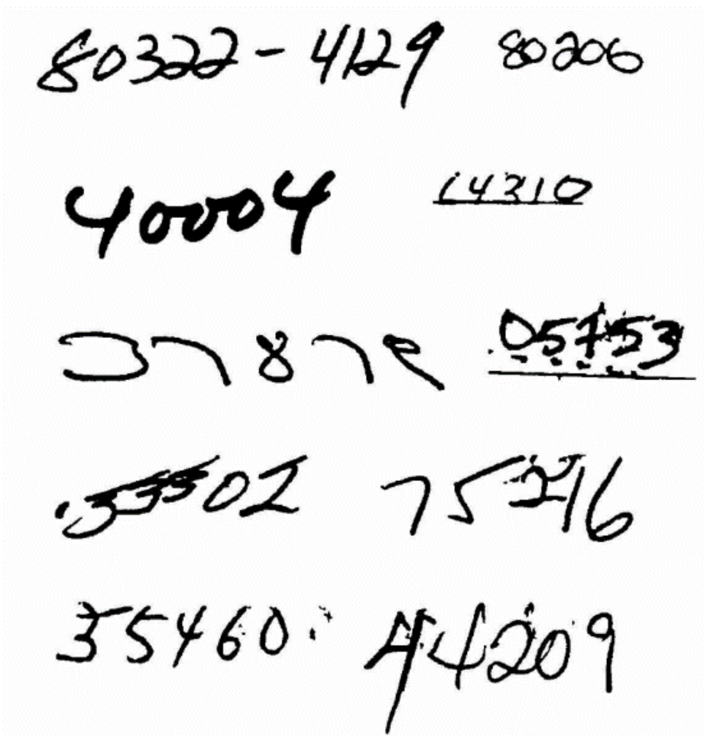
Deep Learning: Hype or Hope?

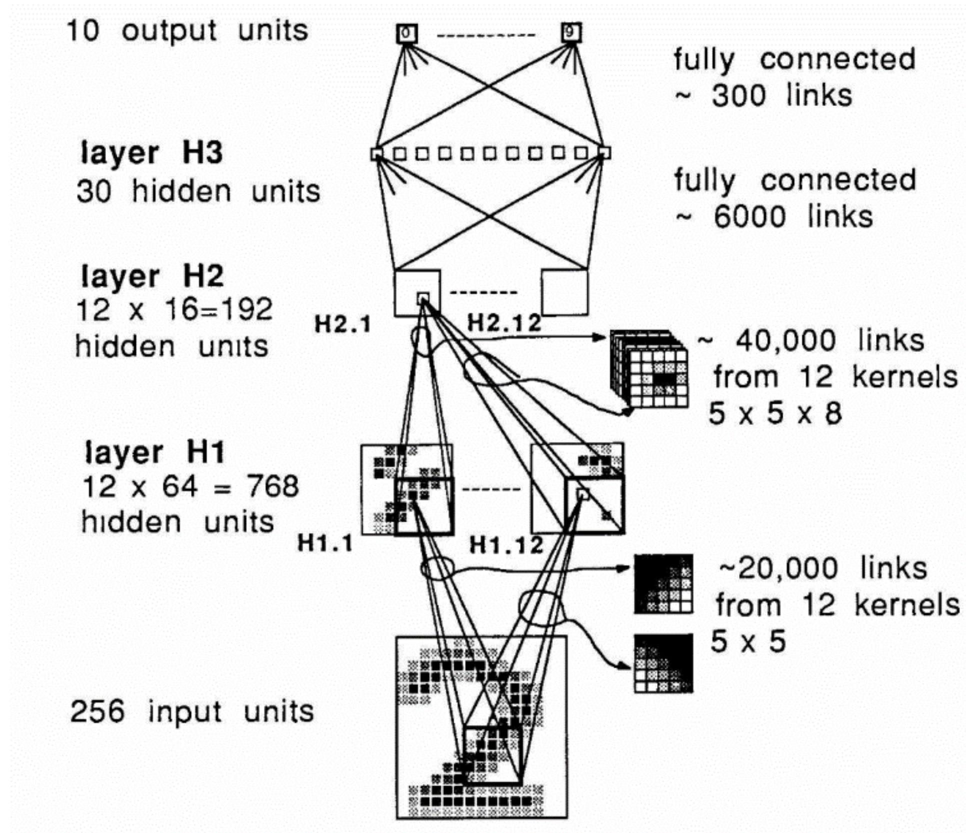
Hype: (n) “extravagant or intensive publicity or promotion”

Hope: (n) “expectation of fulfillment or success”

Milestones: Digit Recognition

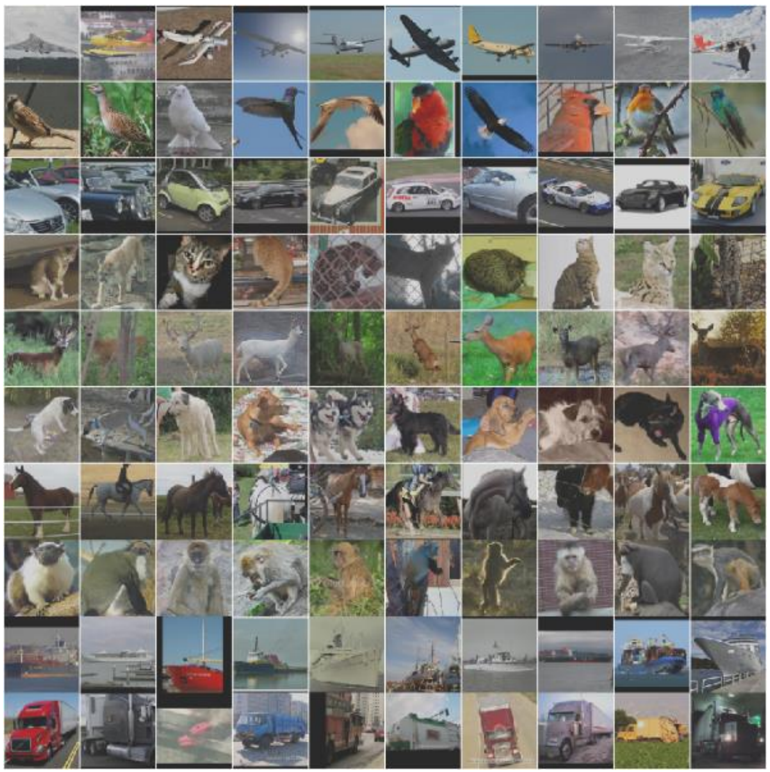
LeNet 1989: recognize zip codes, Yann Lecun, Bernhard Boser and others, ran live in US postal service



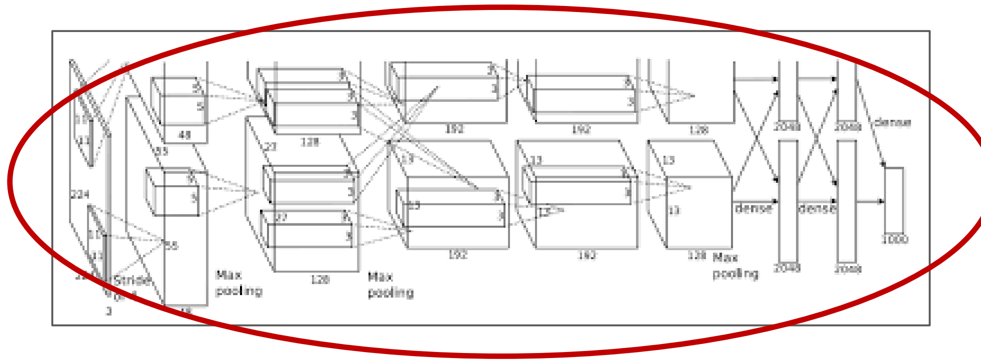


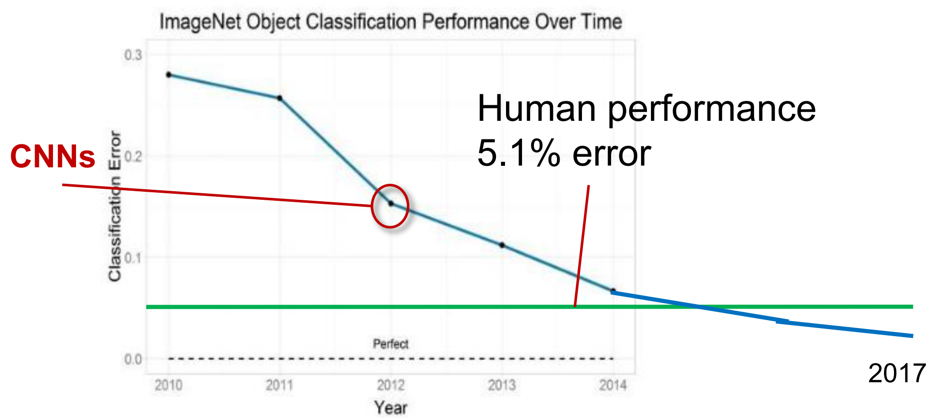
Milestones: Image Classification

Convolutional NNs: AlexNet (2012): trained on 200 GB of ImageNet Data

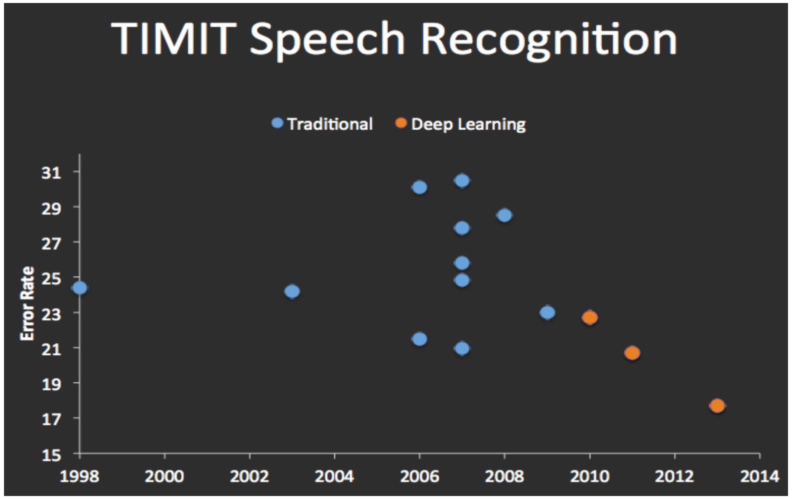


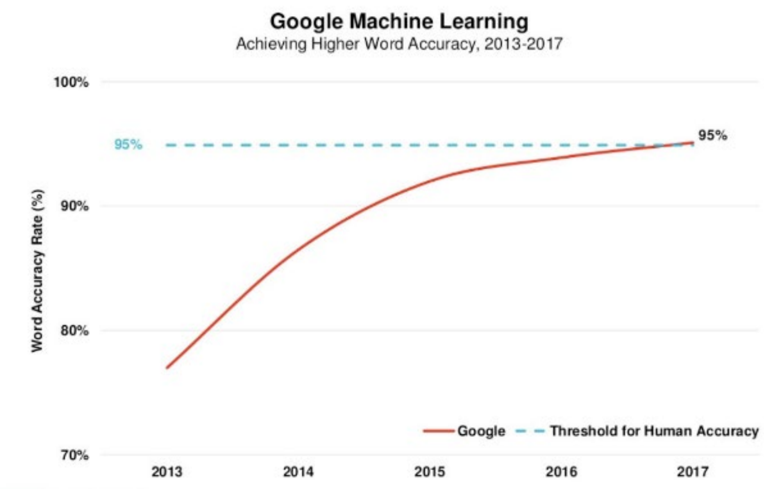
Almost a LeNet !





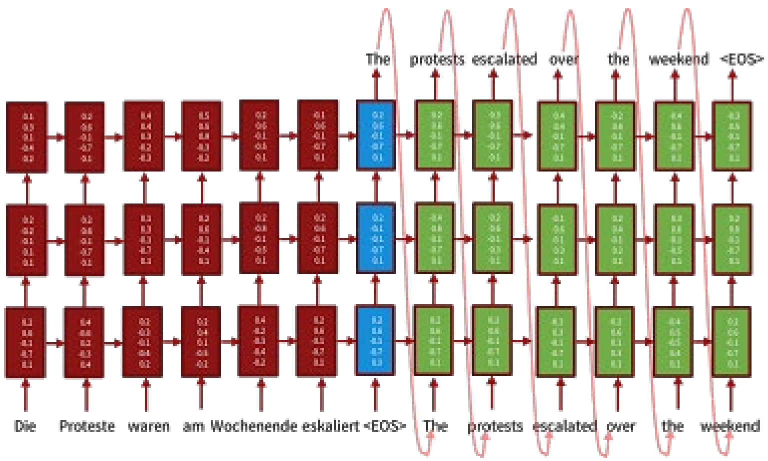
Milestones: Speech Recognition

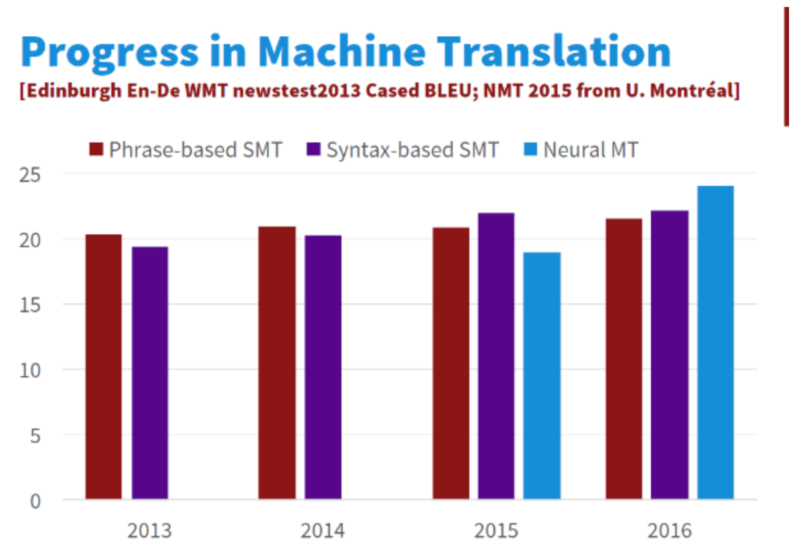




Milestones: Language Translation

Sequence-to-sequence models with recurrent Nets and attention:





From [Sennrich 2016, http://www.meta-net.eu/events/meta-forum-2016/slides/09_sennrich.pdf]

Source Luong, Cho, Manning ACL Tutorial 2016.

Neural Text Processing 2017

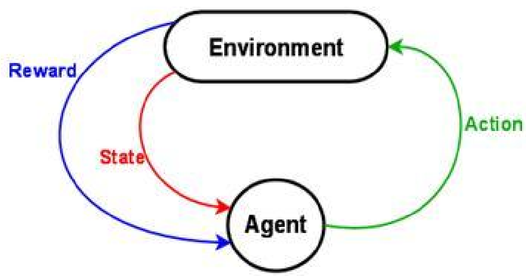
State of the art results on NLP application-level tasks

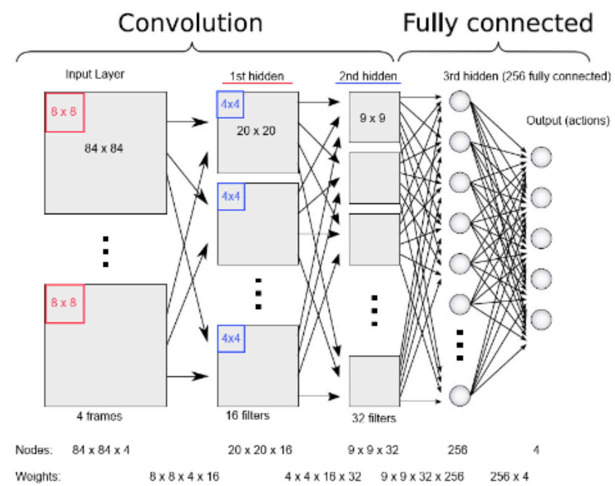
Task	Test set	Metric	Best non-neural	Best neural	Source
Machine Translation	Enu-deu newstest16	BLEU	31.4	34.8	http://matrix.statmt.org
	Deu-enu newstest16	BLEU	35.9	39.9	http://matrix.statmt.org
Sentiment Analysis	Stanford sentiment bank	5-class Accuracy	71.0	80.7	Socher+ 13
Question Answering	WebQuestions test set	F1	39.9	52.5	Yih+ 15
Entity Linking	Bing Query Entity Linking set	AUC	72.3	78.2	Gao+ 14b
Image Captioning	COCO 2015 challenge	Turing test pass%	25.5	32.2	Fang+ 15
Sentence compression	Google 10K dataset	F1	0.75	0.82	Fillipova+ 15
Response Generation	Sordoni dataset	BLEU-4	3.98	5.82	Li+ 16a

[Menezes & Dolan 17]

Milestones: Deep Reinforcement Learning

In 2013, Deep Mind's arcade player bests human expert on six Atari Games. Acquired by Google in 2014,.





In 2016, Deep Mind's alphaGo defeats former world champion Lee Sedol

Milestones: Deep Reinforcement Learning

2017: AlphaGo defeats world champion Ke Jie

Commentators noted that Ke appeared to borrow moves from AlphaGo 2016

But Ke noted that “AlphaGo is improving too fast” and is “a different player from last year”

Risks

2018: OpenAI's Dota agent achieve top 99.95% performance against Human opponents.

AI Risk/Impacts research is very active now, e.g. at DeepMind, and Berkeley's CHAI project.

Deep Learning: Is it Hype or Hope?

Deep Learning: Is it Hype or Hope?

Yes !

Critiques

How smart is today's artificial intelligence?

Today's AI is super impressive, but it's not intelligent.

By Joss Fong | joss@vox.com | Dec 19, 2017, 9:40am EST

INFOWORLD TECH WATCH

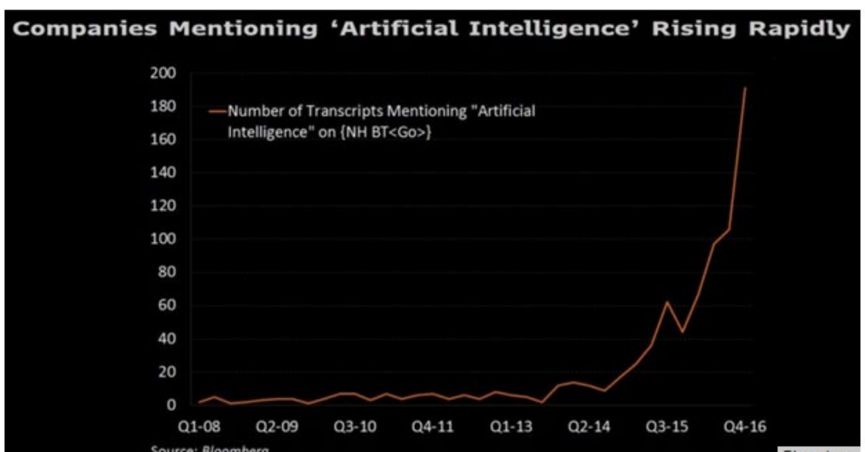
By Matt Asay, InfoWorld | MAR 3, 2017

About

Informed news analysis every weekday

Artificially inflated: It's time to call BS on AI

We may have hit peak ludicrous mode for AI, flailing in a tsunami of AI-washing



Is AI Overhyped?

Forbes Technology Council

Elite CIOs, CTOs & execs offer firsthand insights on tech & business. FULL BIO

Opinions expressed by Forbes Contributors are their own.

POST WRITTEN BY

Ken Weiner

Critiques

How smart is today's artificial intelligence?

Today's AI is super impressive, but it's not intelligent.

By Joss Fong | joss@vox.com | Dec 19, 2017, 9:40am EST

INFOWORLD TECH WATCH

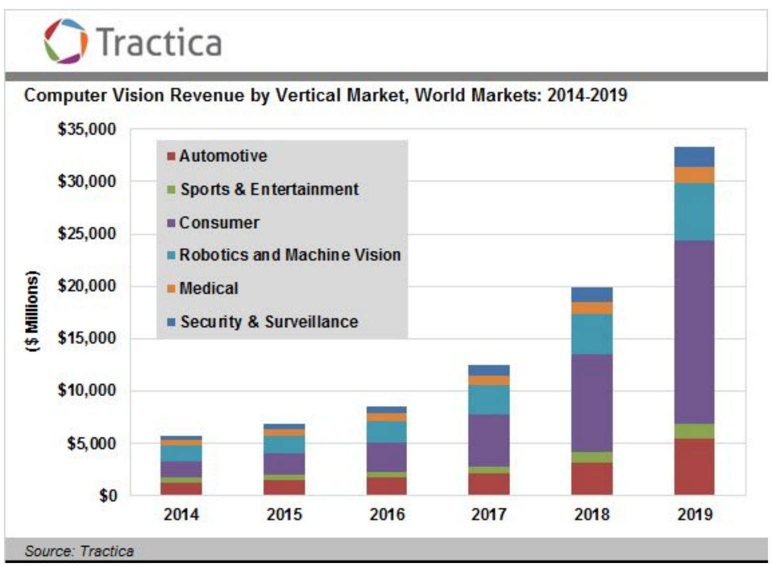
By Matt Asay, InfoWorld | MAR 3, 2017

About

Informed news analysis every weekday

Artificially inflated: It's time to call BS on AI

We may have hit peak ludicrous mode for AI, flailing in a tsunami of AI-washing



Is AI Overhyped?

Forbes Technology Council

Elite CIOs, CTOs & execs offer firsthand insights on tech & business. FULL BIO

Opinions expressed by Forbes Contributors are their own.

POST WRITTEN BY

Ken Weiner

Setting Expectations … Badly

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966

Vision Memo. No. 100.

<u>THE SUMMER VISION PROJECT</u>

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Critiques:

(Some) AI Scientists made over-optimistic predictions about AI system capabilities in the 80s and 90s leading to an “AI winter.” So optimistic predictions about AI systems today should be ignored…

- The assertion relies on faulty logic:
  - AI systems → intrinsically poor performance
  - AI researcher → always over-estimate

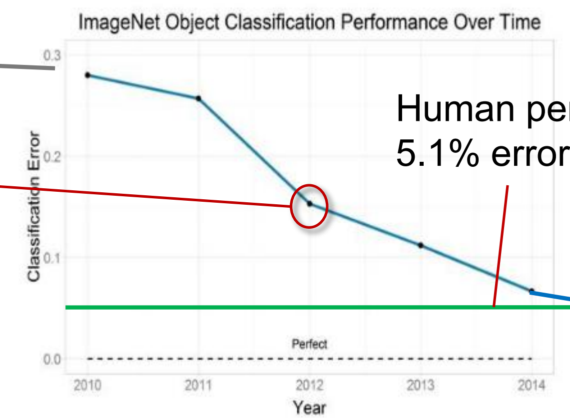
AI systems are still not “generally intelligent,” … even if they are awfully good at an awful lot of things.

- As a concept “general intelligence” has been the bane of AI research.
- Progress in AI has recently been dictated mostly by economic importance, not “AI-hardness”. The “AI-hard” list shrinks every year:
Jeopardy, Go, Science Exams, Face recognition, Translation,…

Data on Classical AI vs. Deep Learning:

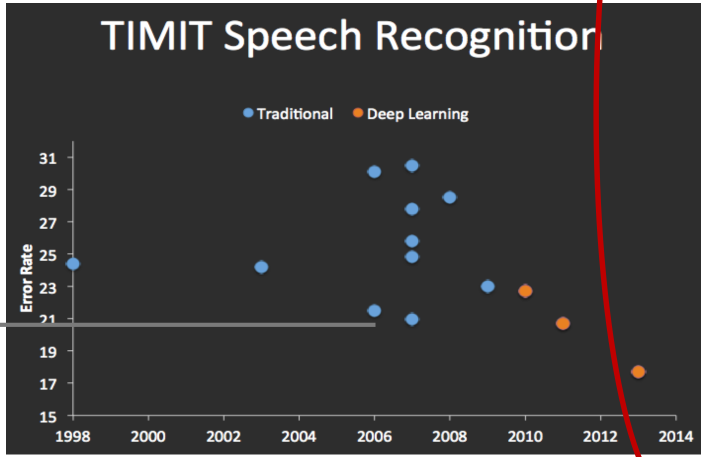
Performance floor

CNNs



Human performance
5.1% error

2017



Performance floor

Where’s the floor ?

Opportunities:

Of course!

- Science, engineering, entertainment, education, communication, organization, recreation, medicine, driving, games (real and virtual), transportation, commerce, e-trading, name-your-topic…

Risks:

Yes!

- Economic: displacing jobs

- Existential: security, systems running amok

Hawking, Musk, Gates have been highlighting the risks of new AI technologies.

Learning about Deep Neural Networks

Yann Lecun (Facebook research head, DNN pioneer) quote: DNNs require:
“an interplay between intuitive insights, theoretical modeling, practical implementations, empirical studies, and scientific analyses”

i.e. there isn’t a single framework or core set of principles to explain everything (c.f. graphical models for machine learning).

We try to cover the ground in Lecun’s quote.

This Course (please interrupt with questions)

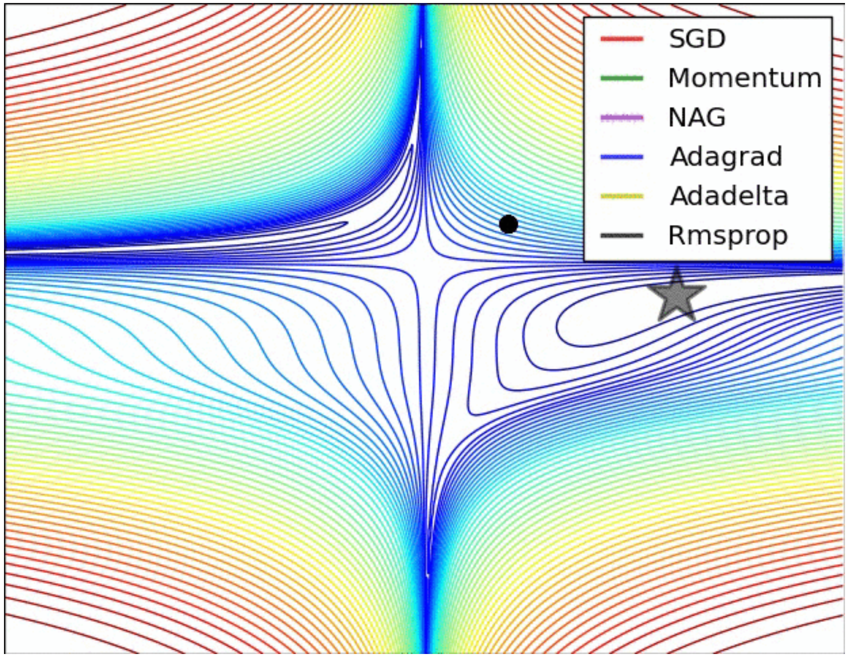
Goals:

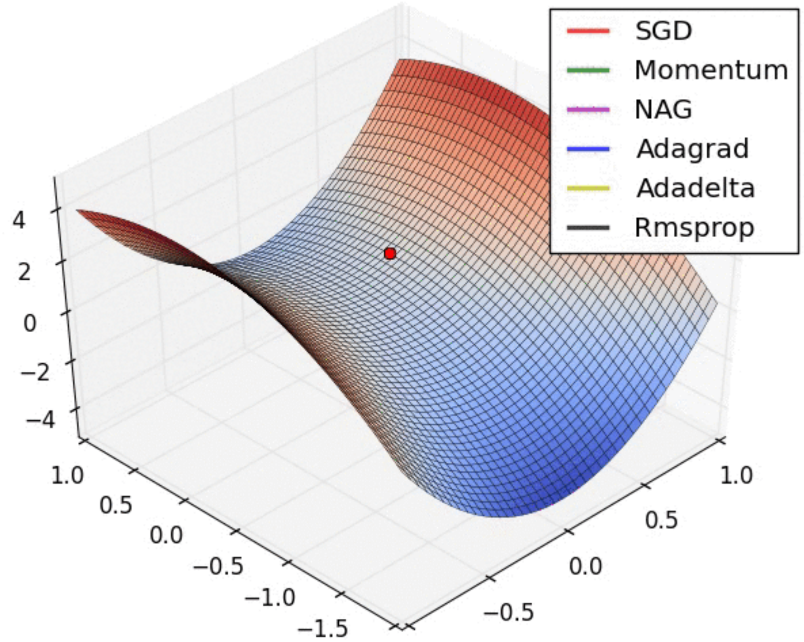
- Introduce deep learning to a broad audience.
- Review principles, techniques and visualization for understanding deep networks.
- Develop skill at designing networks for applications.

Materials:

- Book(s)
- Notes
- Lectures

The role of Animation





From A. Karpathy's cs231n notes.

This Course: bCourses page

- The course will run through bCourses.
- The bCourse Home page is the master link to everything.
- The “syllabus” page contains only assignment deadlines, don’t use it for other purposes.

bCourses Home page has the schedule

Date	Lecture Topic	Reading	Assignments/Section Notes
Tu 1/21	Introduction, Course Overview, Brief history of deep networks. pptx or pdf	Introduction from Deep Learning	
Th 1/23	Machine learning concepts: Loss and Risk, Discriminative models, Linear and Logistic Regression. pptx or pdf	Review: sections 1.1-1.3 and 6.6-6.9 from the CS189 book (skip KL-div). Do Python/Numpy tutorial if needed.	Section 1 notes pdf
Tu 1/28	Bias-Variance tradeoff, Regularization, SVMs, Multiclass classification, Softmax. Cross-validation. pptx and pdf	sections 1.4-1.6, 6.10-6.11, from the CS189 book (skip Tikhonov)	Assignment 1 out
Th 1/30	Optimization, Stochastic Gradient Descent. pptx and pdf	Chapter 8 of Deep Learning Optimization Notes	Section 2 notes pdf
Tu 2/4	Backpropagation, Convolutional Networks. pptx and pdf	Backpropagation Notes Convnet notes	
Th 2/6	CNN examples, Activation functions, Initialization. pptx and pdf	Convnet notes Training Neural Networks 1 Training Neural Networks 2	Project Proposal out Section 3 notes pdf
Tu 2/11	Training: Batch normalization, Dropout, Ensembles, Hyperparameter tuning. pptx and pdf	Training Neural Networks 2 Training Neural Networks 3	
Th 2/13	TBD		
			Assignment 1 due 11pm Assignment 2 out
Tu 2/18	Object Detection and Segmentation. pptx and pdf	Introduction to Object Detection (All sections *except* ParseNet, PSPNet, DeepLab, PANet and EncNet)	Project proposal due Practice MT1 Older MT1

bCourses Schedule (home page)

Intro., ML review

Computer Vision, General DNN principles

Natural language, Generative and Adversarial Nets

Imitation and reinforcement learning

Schedule

Date	Lecture Topic	Reading	Assignments/Section Notes
Tu 1/21	Introduction, Course Overview, Brief history of deep networks. pptx or pdf	Introduction from Deep Learning	
Th 1/23	Machine learning concepts: Loss and Risk, Discriminative models, Linear and Logistic Regression. pptx or pdf	Review: sections 1.1-1.3 and 6.6-6.9 from the CS189 book (skip KL-div). Do Python/Numpy tutorial if needed.	Section 1 notes pdf
Tu 1/28	Bias-Variance tradeoff, Regularization, SVMs, Multiclass classification, Softmax. Cross-validation. pptx and pdf	sections 1.4-1.6, 6.10-6.11, from the CS189 book (skip Tikhonov)	Assignment 1 out
Th 1/30	Optimization, Stochastic Gradient Descent. pptx and pdf	Chapter 8 of Deep Learning Optimization Notes	Section 2 notes pdf
Tu 2/4	Backpropagation, Convolutional Networks. pptx and pdf	Backpropagation Notes Convnet notes	
Th 2/6	CNN examples, Activation functions, Initialization. pptx and pdf	Convnet notes Training Neural Networks 1 Training Neural Networks 2	Project Proposal out Section 3 notes pdf
Tu 2/11	Training: Batch normalization, Dropout, Ensembles, Hyperparameter tuning. pptx and pdf	Training Neural Networks 2 Training Neural Networks 3	
Th 2/13	TBD		
			Assignment 1 due 11pm Assignment 2 out
Tu 2/18	Object Detection and Segmentation. pptx and pdf	Introduction to Object Detection (All sections *except* ParseNet, PSPNet, DeepLab, PANet and EncNet) Introduction to Semantic Segmentation Methods	Project proposal due Practice MT1 Older MT1
Th 2/20	Recurrent Networks, LSTMs, applications. pptx and pdf	RNN chapter from Deep Learning Understanding LSTM Networks	Section 4 notes pdf
Tu 2/25	Visualizing Deep Networks. pptx and pdf	Understanding Neural Networks Through Deep Visualization Feature Visualization	Practice Mid-term Solutions
TBD	Midterm 1		
Th 2/27	Semantic Models for Text. pptx and pdf	WordToVec Skip-Thought Vectors	Project Checkpoint 1 due week of 3/11/19
Tu 3/3	Attention Networks. pptx and pdf	Recurrent Models of Visual Attention	Assignment 2 due 11pm Assignment 3 out Section 6 Notes pdf
Th 3/5	Natural Language Translation. pptx and pdf	NMT by Jointly Learning to Align and Translate Attention is All You Need The Illustrated Transformer	
Tu 3/10	Text Question Answering. pptx and pdf	End-to-End Memory Networks QANet	Section 7 Notes pdf
Th 3/12	Neural Dialog Systems. pptx and pdf	Learning End-to-End Goal Directed Dialog BERT: Pre-training of Deep Bidirectional Transformers	
Tu 3/17	Adversarial and Fooling Networks. pptx and pdf	Adversarial Examples Transferability/blackbox attacks Physical perturbation	Practice Midterm 2 Section 8 Notes pdf
Th 3/19	Generative Models pptx and pdf	Variational Auto-Encoders Autoregressive Models Image Transformer	Assignment 3 due 11pm Assignment 4 out
TBD	MT2 review session		
Tu 3/31	Generative Adversarial Networks pptx pdf	Generative Adversarial Networks DCGAN Paper Wasserstein GAN Blog Post (From GAN to WGAN)	Section 9 Notes pdf
TBD	Midterm 2		
Th 4/2	Imitation Learning pdf	End-to-End Learning for Self-Driving Cars DAGGER Adversarial Imitation Learning	No Sections
Tu 4/7	Reinforcement Learning: Policy Gradients pdf	Policy Gradient Methods, chapter 13 of "Reinforcement Learning"	
Th 4/9	Reinforcement Learning: Value-based methods pdf	DQN paper (Nature) Asynchronous Methods for Deep RL Double DQN	Section 10 Notes pdf
Tu 4/14	Exploration pdf	Count-Based Exploration and Intrinsic Motivation Curiosity-driven Exploration Infobot: Information Bottleneck for RL	
TBD			Assignment 4 due 11pm
Th 4/16	Learning to Learn pdf	Learning to Learn blog post Matching Networks for One-Shot Learning Model-Agnostic Meta-Learning	Section 11 Notes pdf
Tu 4/21	Playing Games pptx and pdf	A Survey of Monte-Carlo Tree Search Methods sections 1-3 Mastering the Game of Go without Human Knowledge	
			Project Poster due 11pm
	Final project poster session		Project Poster
			Project Report due 11pm

This Course

Work:

- Class Participation: 10%
- 2 Midterms: 30%
- Final Project (in groups): 30%
- 4 Assignments : 30%

Audience: primarily EECS undergrads and grads.

Discussion Sections

No discussions this week, they start next Week.

Class Format Rationale: L and W sections

Empirically, in previous semesters, we’ve found that students either attend almost all lectures, or almost none.

Standard lecture-format courses waste resources in that situation (huge mostly-empty classrooms).

The experience is mostly-online, but content is optimized for in-class delivery.

Instructor time is dominated with logistics rather than pedagogy.

Class Format Rationale: L and W sections

L sections let us use classroom space efficiently, and encourage interaction.

Content can be optimized for online delivery.

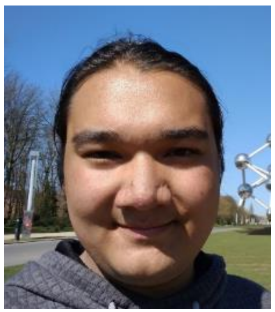
Course manager assumes more of the course logistics, freeing the instructor to focus on content and pedagogy.

Course Staff

Prof: John Canny

Course Manager:
Michael-David Sasson

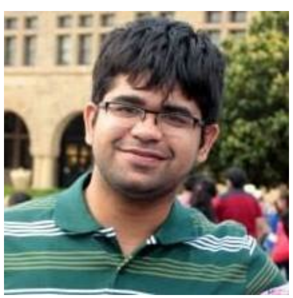
GSIs:



David Chan

Haozhi Qi

Philippe Laban



Aravind Srinivas

Roshan Rao

Forrest Huang

Prerequisites

- Knowledge of calculus and linear algebra, Math 53/54 or equivalent.
- Probability and Statistics, CS70 or Stat 134. CS70 is bare minimum preparation, a stat course is better.
- Machine Learning: CS189, strongly encouraged but not required.
- Programming, CS61B or equivalent. Assignments will mostly use Python.

Logistics

- Course Number: CS L182/W182/282A, SP 2020
- On bCourses, most material is publicly readable.
- Instructor: John Canny lastname@berkeley.edu
- Time: TuTh 9:30am – 11:00am
- Location: 306 Soda
- Discussion: Join Piazza for announcements and to ask questions about the course
- Section times, staff office hours are on the main page
- Webcasts (video) will appear on our youtube channel.
- We will also post augmented powerpoint slides (with audio and Q/A) after lecture.

Logistics

- iClickers ?

Course Project

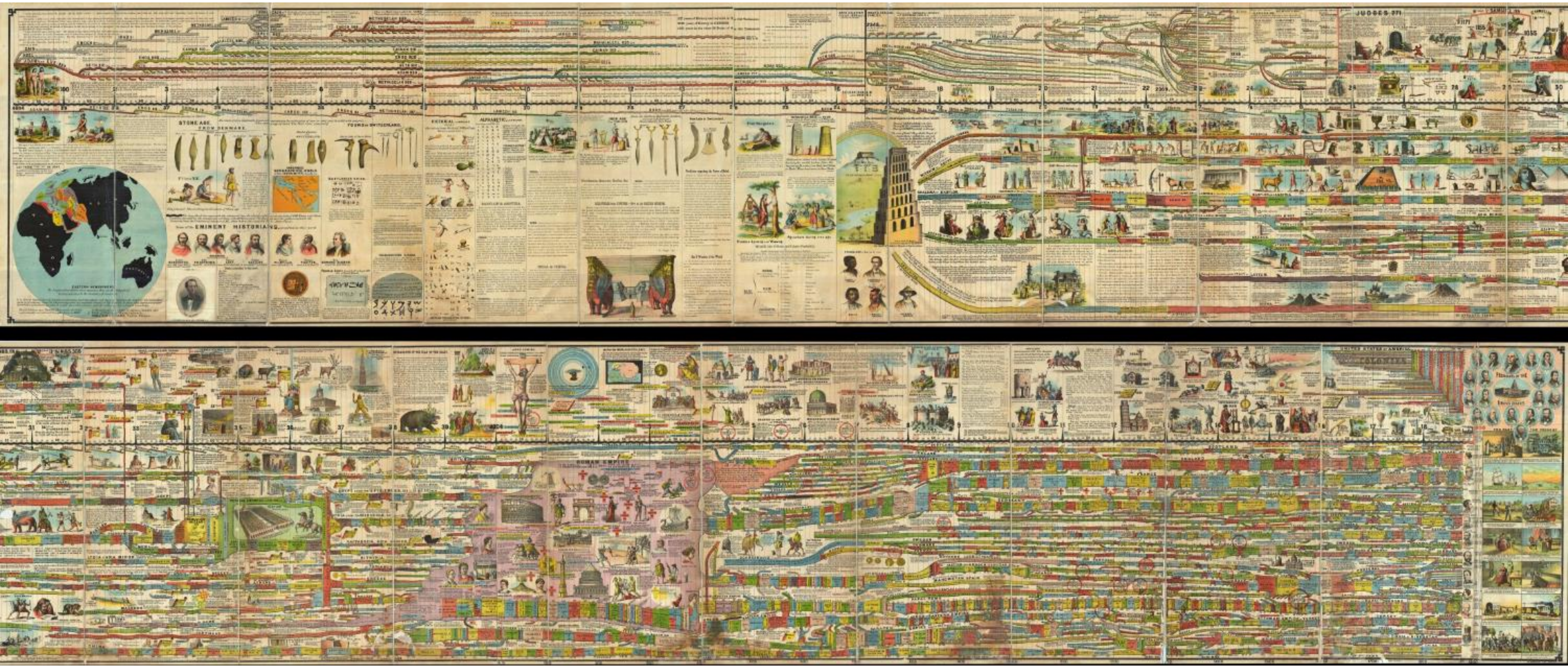
- Will consume about 2/3 of the semester.
- In teams of 3-4.
- Can be combined with other course projects
- We encourage “open-source” projects that can be archived.
- You will “check-in” with the GSIs several times during the semester.
- Final video and report due at the end of the semester.
- Difference from last year: CS182 projects will be limited to a set of Kaggle challenges. CS 282A projects will be open-ended.

Questions ?

Coming up: Some rationale for deep neural networks…

Some History

- Reading: the Deep Learning Book, Introduction



Phases of Neural Network Research

- 1940s-1960s: Cybernetics: Brain like electronic systems, morphed into modern control theory and signal processing.
- 1960s-1980s: **D**igital computers, automata theory, computational complexity theory: simple shallow circuits are very limited in what they can represent…
- 1980s-1990s: Connectionism: complex, non-linear networks, back-propagation.
- 1990s-2010s: Computational learning theory, graphical models: Learning is computationally hard, simple shallow circuits are very limited in what they can learn…
- 2006→: Deep learning: End-to-end training, large datasets, explosion in applications.

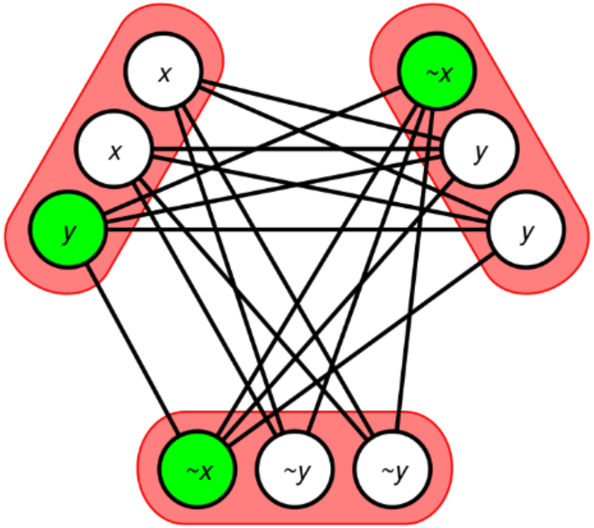
Citations of the "LeNet" paper

- Recall the LeNet was a modern visual classification network that recognized digits for zip codes. Its citations look like this:

- The 2000s were a golden age for machine learning, and marked the ascent of graphical models. But not so for neural networks.

Why the success of DNNs is surprising

- From both complexity and learning theory perspectives, simple neural networks are very limited.
  - Can’t compute parity with a small network.
  - NP-Hard to learn “simple” functions like 3SAT formulae, and i.e. training a DNN is NP-hard.



Why the success of DNNs is surprising

- The most successful DNN training algorithm is a version of gradient descent which will only find local optima. In other words, it's a greedy algorithm.

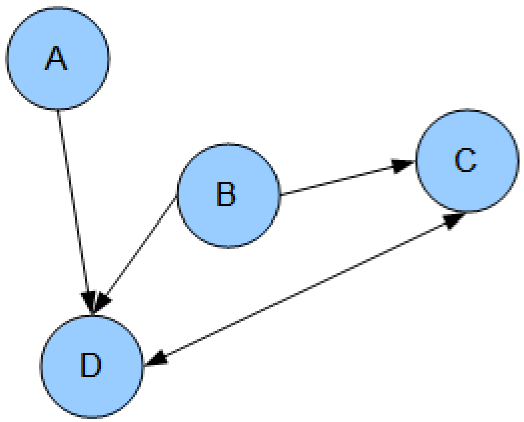
- Greedy algorithms are even more limited in what they can represent and how well they learn.

- If a problem has a greedy solution, its regarded as an "easy" problem.

Why the success of DNNs is surprising

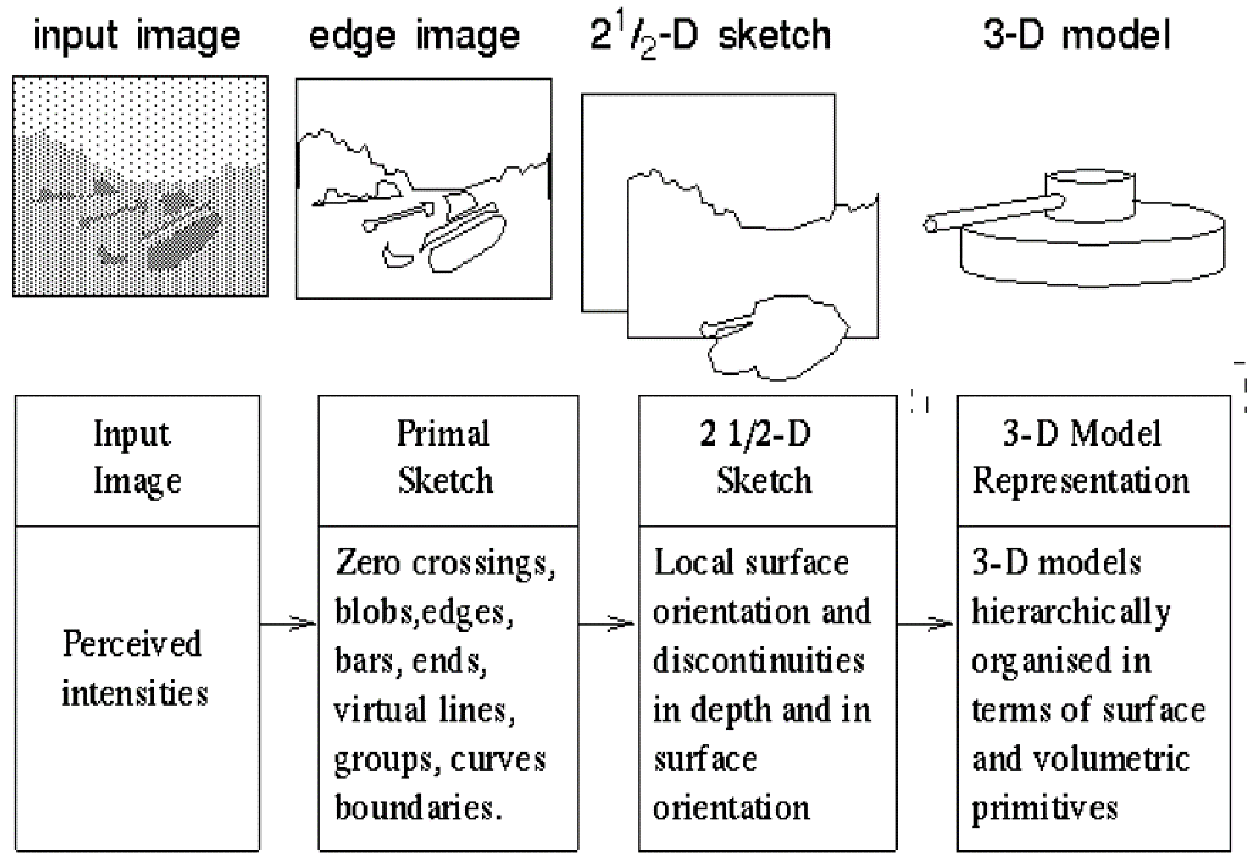
- In graphical models, values in a network represent random variables, and have a clear meaning. The network structure encodes dependency information, i.e. you can represent rich models.

- In a DNN, node activations encode nothing in particular, and the network structure only encodes (trivially) how they derive from each other.



Why the success of DNNs is ~~surprising~~ obvious

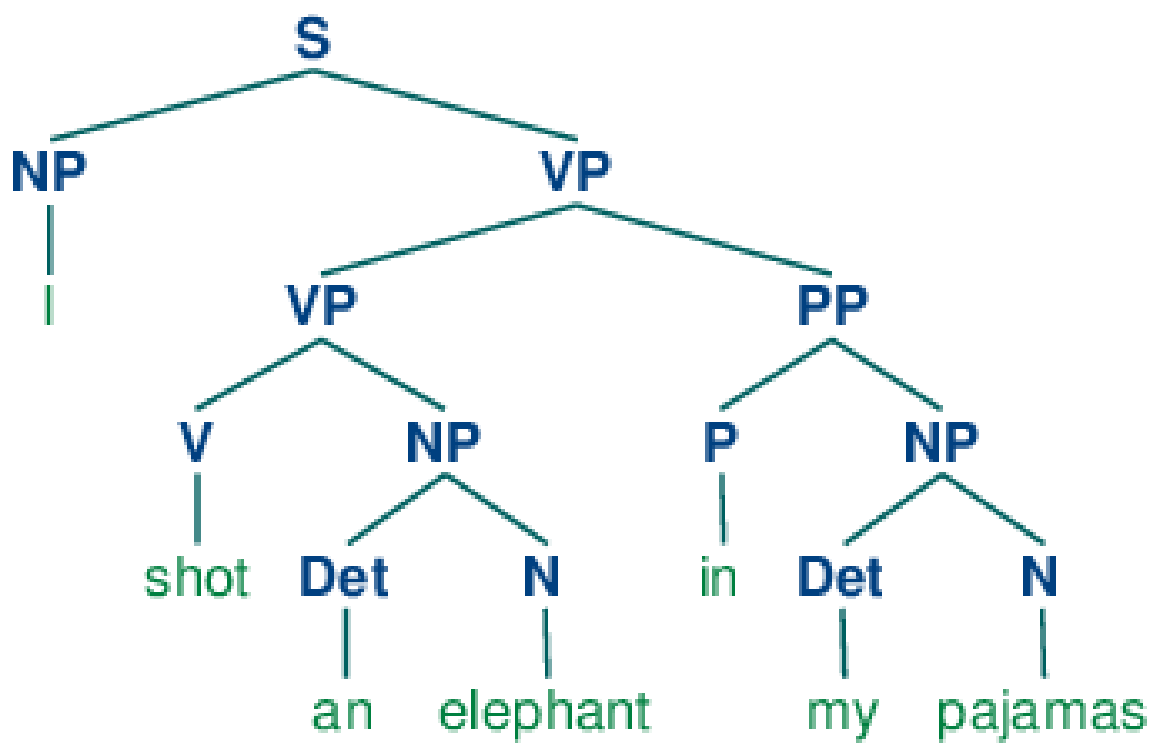
- Hierarchical representations are ubiquitous in AI. Computer vision:



Stages of Visual Representation, David Marr, 1970s

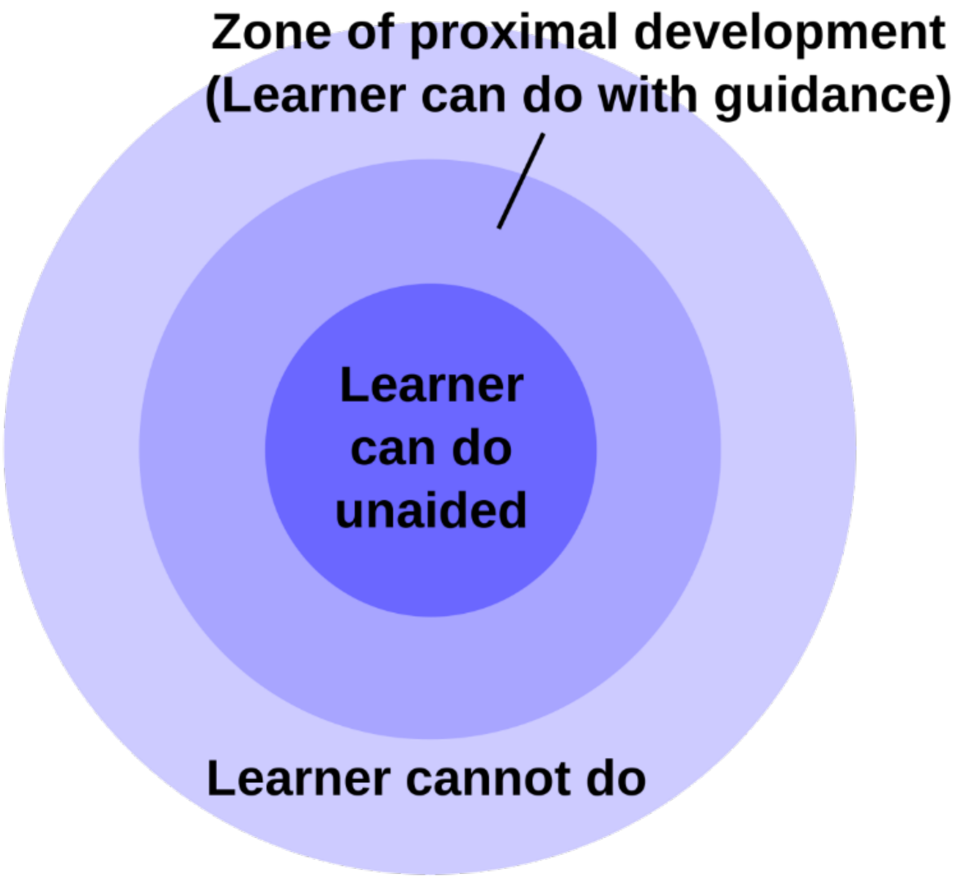
Why the success of DNNs is ~~surprising~~ obvious

- Hierarchies are ubiquitous in natural language:



Why the success of DNNs is ~~surprising~~ obvious

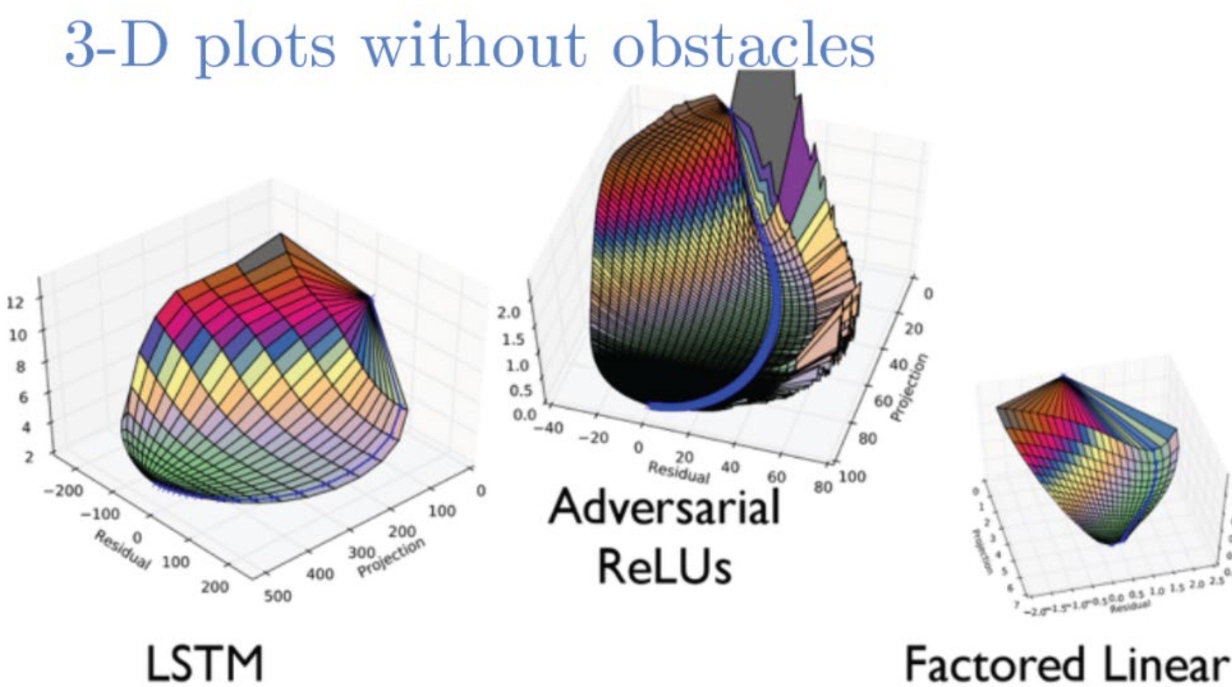
- Human Learning: is deeply layered.



Deep expertise

Why the success of DNNs is ~~surprising~~ obvious

- What about greedy optimization?
- Less obvious, but it looks like many learning problems (e.g. image classification) are actually “easy” i.e. have reliable steepest descent paths to a good model.

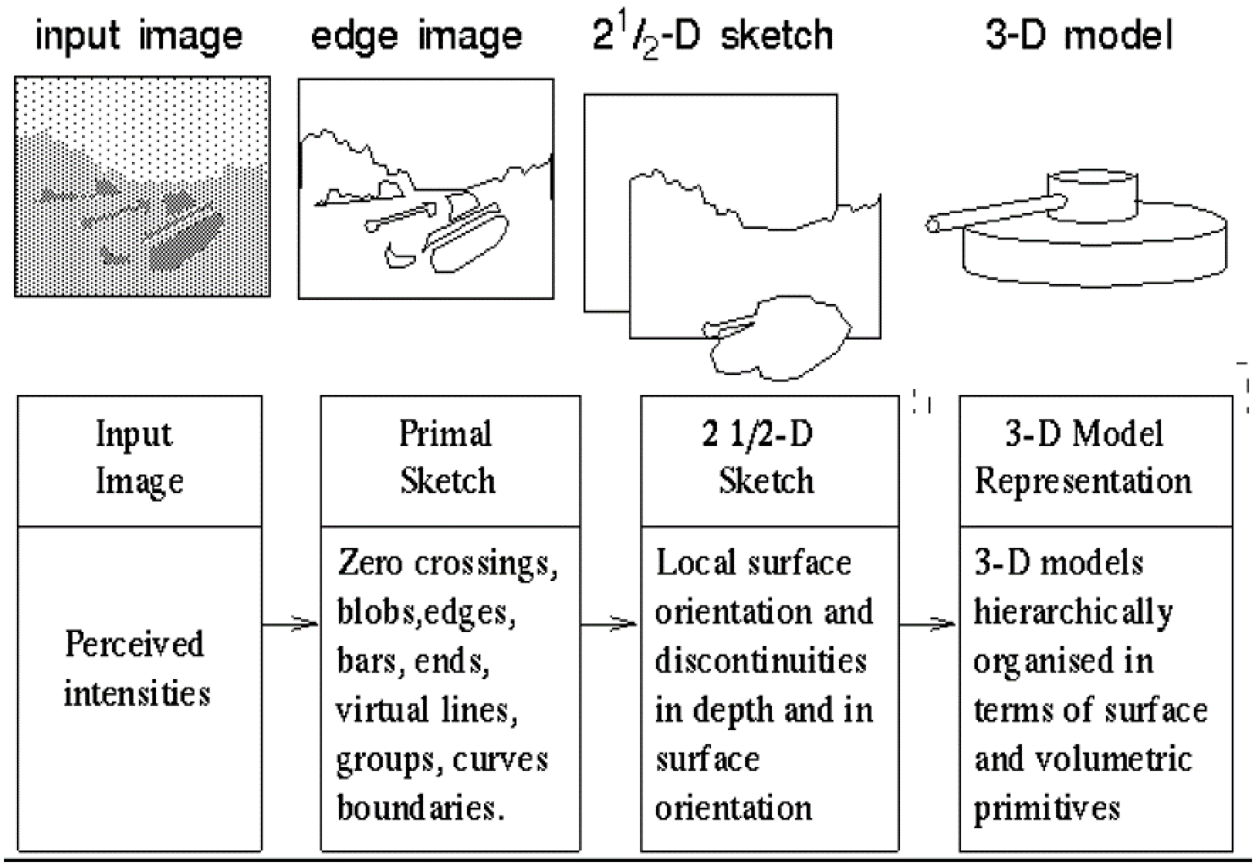


Questions ?

Coming up: Representation Learning

Hierarchical Representations

Hierarchical representations are ubiquitous in AI, but had to be hand-designed. In computer vision again:

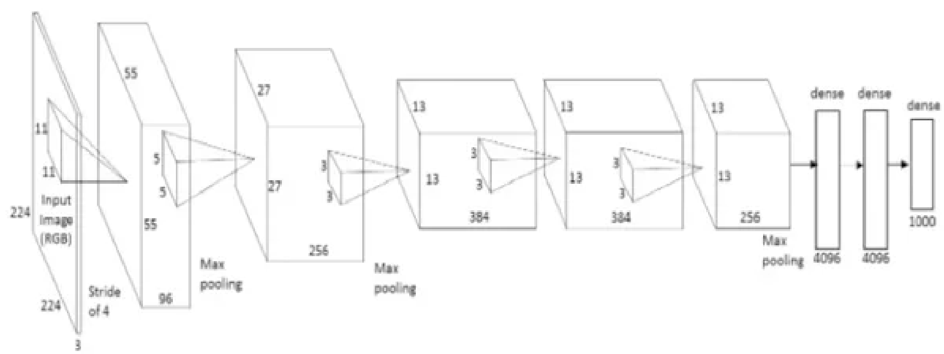


Stages of Visual Representation, David Marr, 1970s

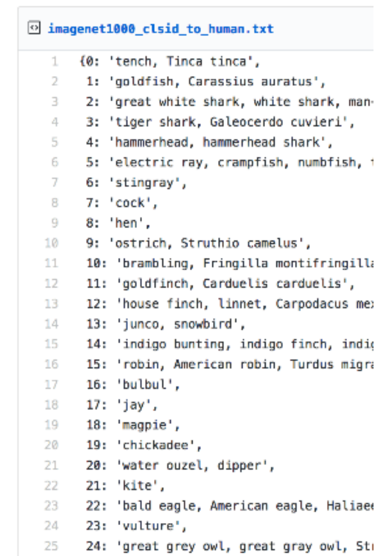
Representation Learning

Learning Layered Representations: Training a deep Imagenet classifier:

1.4 million images



“Alexnet” (Krizhevsky) from 2012



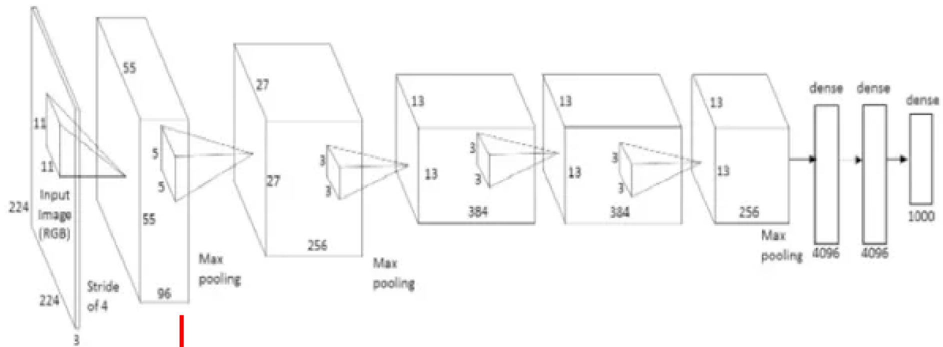
imagenet1000_clsid_to_human.txt

```
{0: 'tench, Tinca tinca',
 1: 'goldfish, Carassius auratus',
 2: 'great white shark, white shark, man-
 3: 'tiger shark, Galeocerdo cuvieri',
 4: 'hammerhead, hammerhead shark',
 5: 'electric ray, crampfish, numbfish, t
 6: 'stingray',
 7: 'cock',
 8: 'hen',
 9: 'ostrich, Struthio camelus',
 10: 'brambling, Fringilla montifringilla
 11: 'goldfinch, Carduelis carduelis',
 12: 'house finch, linnet, Carpodacus mex
 13: 'junco, snowbird',
 14: 'indigo bunting, indigo finch, indig
 15: 'robin, American robin, Turdus migra
 16: 'bulbul',
 17: 'jay',
 18: 'magpie',
 19: 'chickadee',
 20: 'water ouzel, dipper',
 21: 'kite',
 22: 'bald eagle, American eagle, Haliaee
 23: 'vulture',
 24: 'great grey owl, great gray owl, Str
```

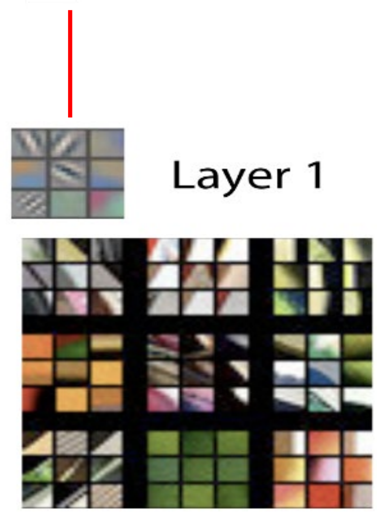
1000 labels

Representation Learning

After a lot of training…



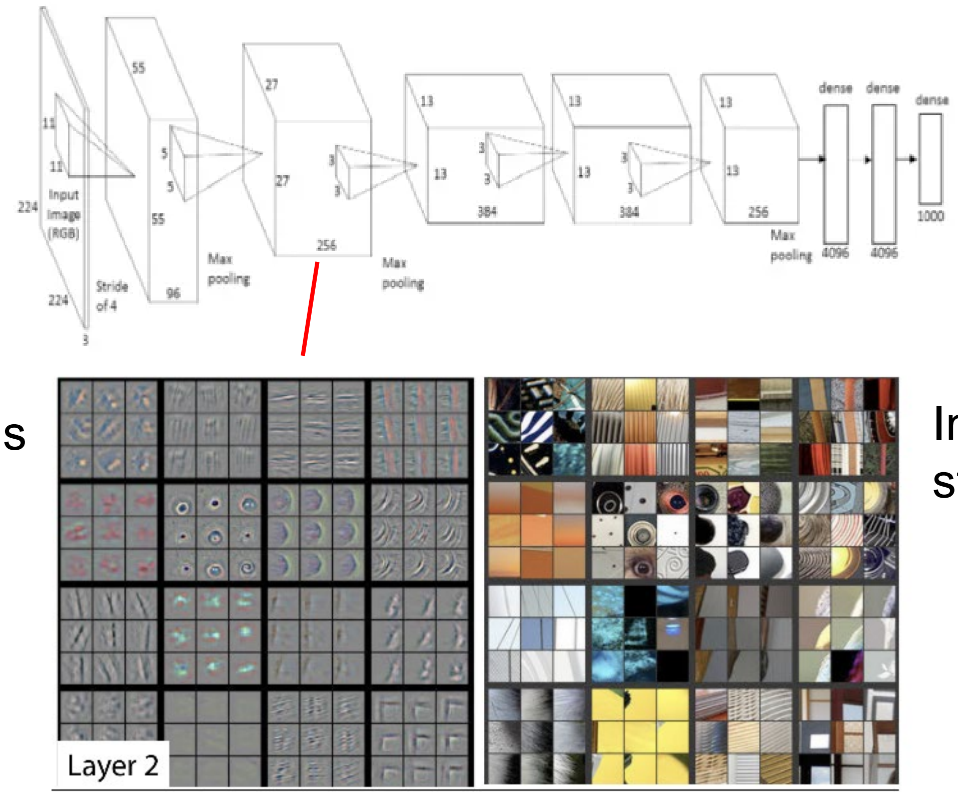
Neural receptive field
gray is zero
black is negative
white is positive



Zeiler and Fergus, “Visualizing and Understanding Convolutional Networks”, 2013

Image patches with strongest responses

Representation Learning



Neural receptive fields
gray is zero
black is negative
white is positive

Image patches with strongest responses

Zeiler and Fergus, “Visualizing and Understanding Convolutional Networks”, 2013

Representation Learning

Neural receptive fields

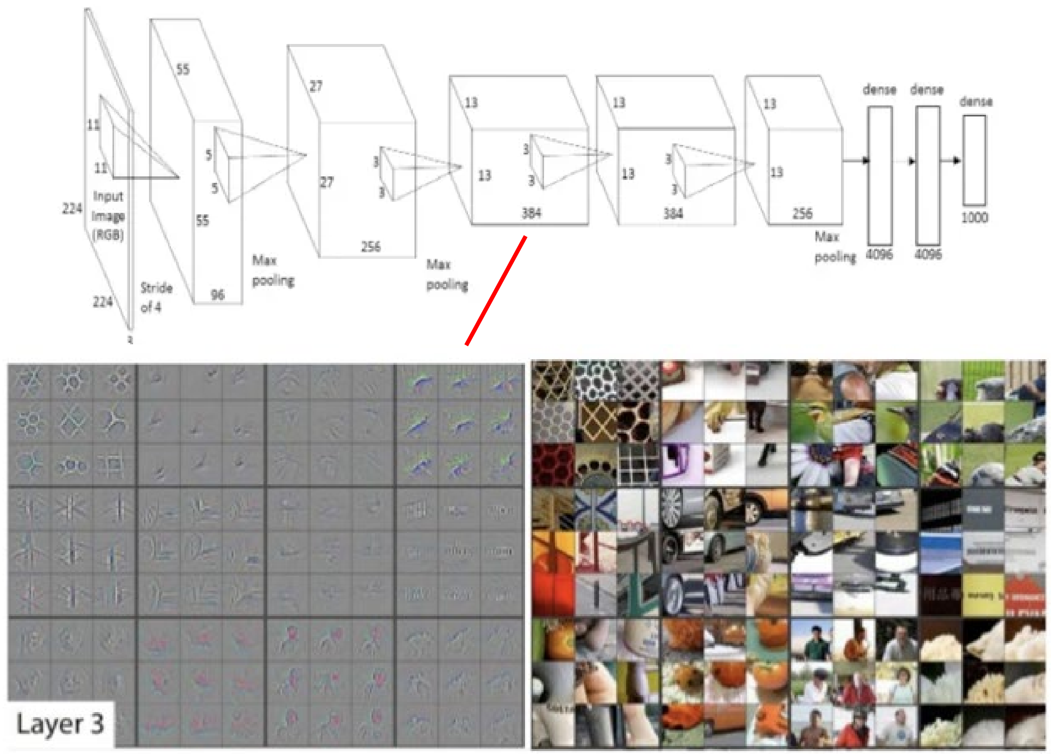
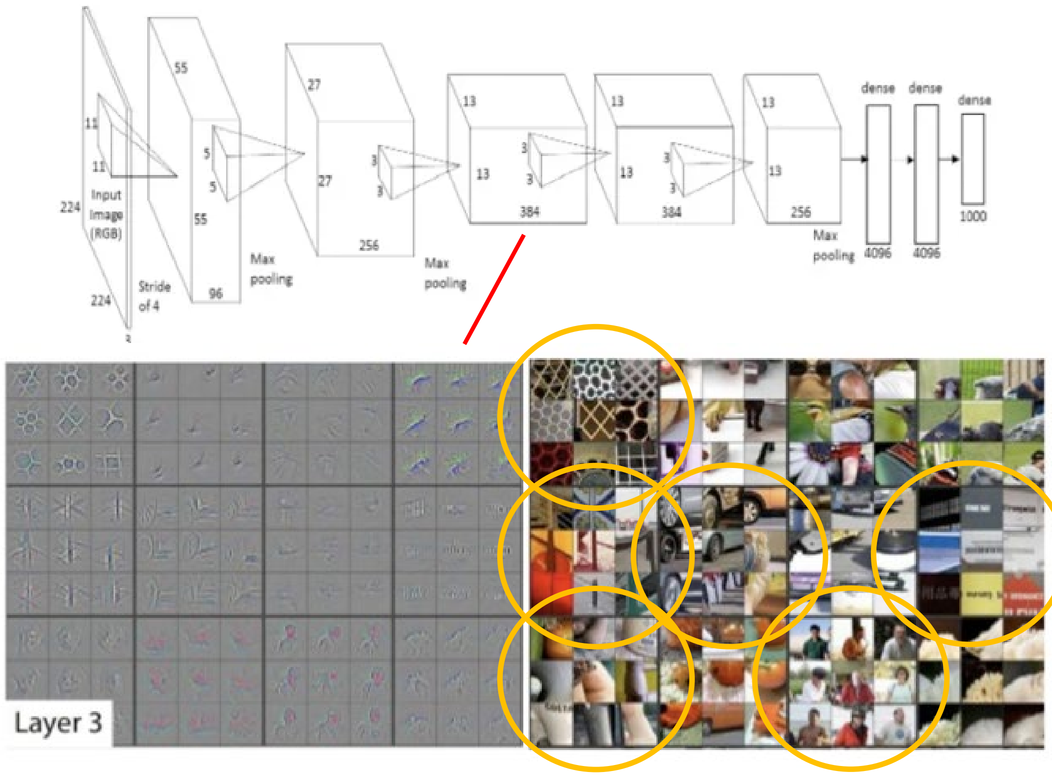


Image patches with strongest responses

Zeiler and Fergus, “Visualizing and Understanding Convolutional Networks”, 2013

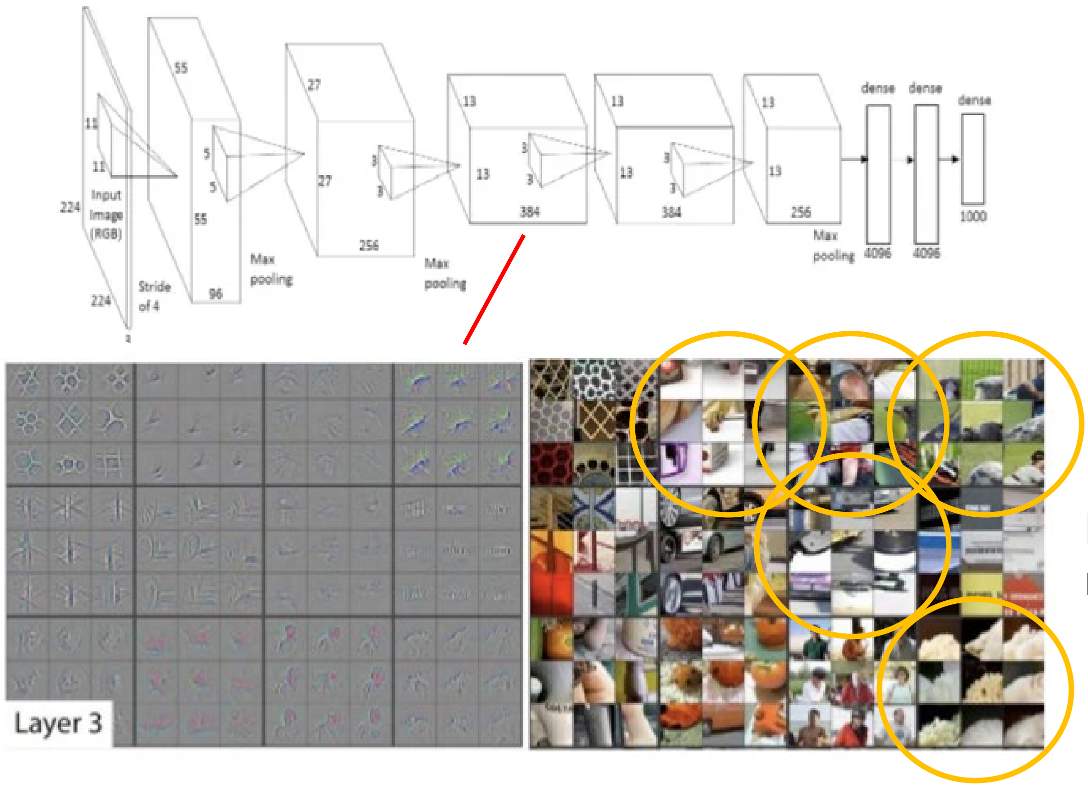
Representation Learning



Interpretable patches

Zeiler and Fergus, “Visualizing and Understanding Convolutional Networks”, 2013

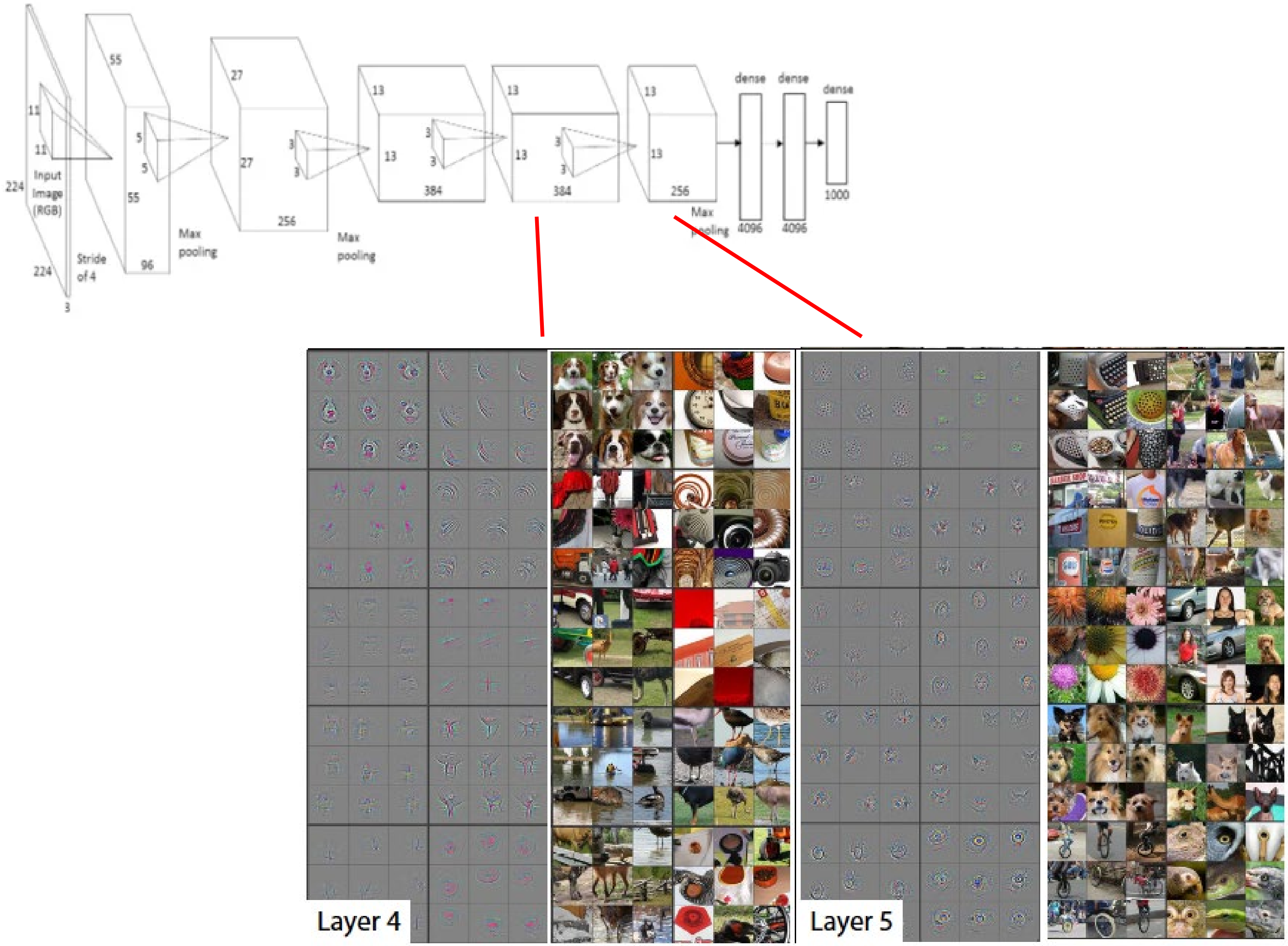
Representation Learning



Not so much…

Zeiler and Fergus, "Visualizing and Understanding Convolutional Networks", 2013

Representation Learning



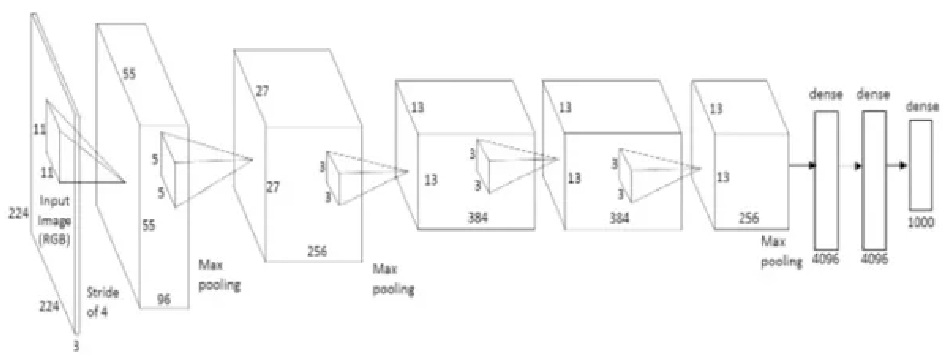
Zeiler and Fergus, “Visualizing and Understanding Convolutional Networks”, 2013

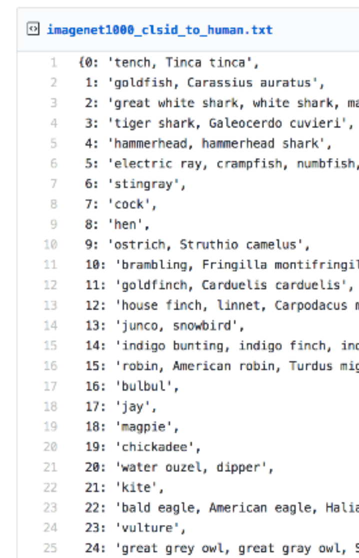
Representation Learning

These inner representations were learned ***only*** from input image/output labels.

Early and intermediate layer weights are typically task-independent.

1.4 million images





1000 labels

Consequences

You can reuse or customize models trained for different task.

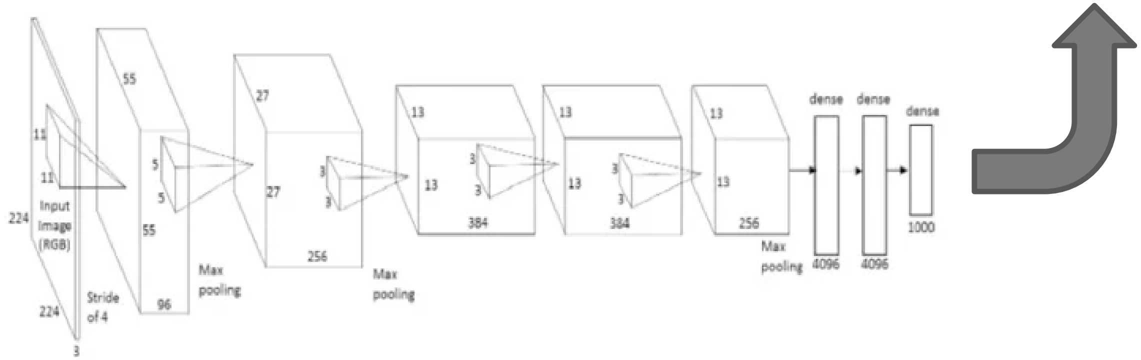
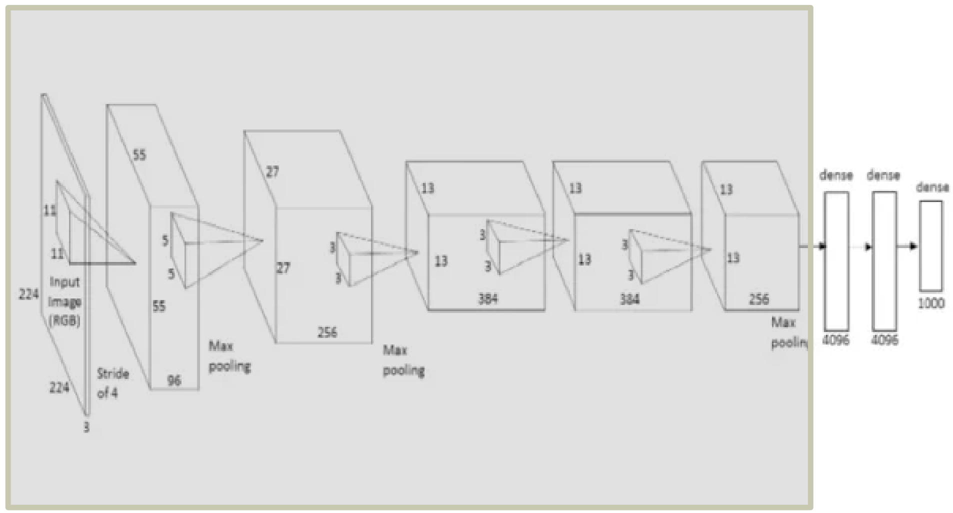


Image
Classification

Reuse/Fine Tuning

You can reuse or customize models trained for different task, with much less data.

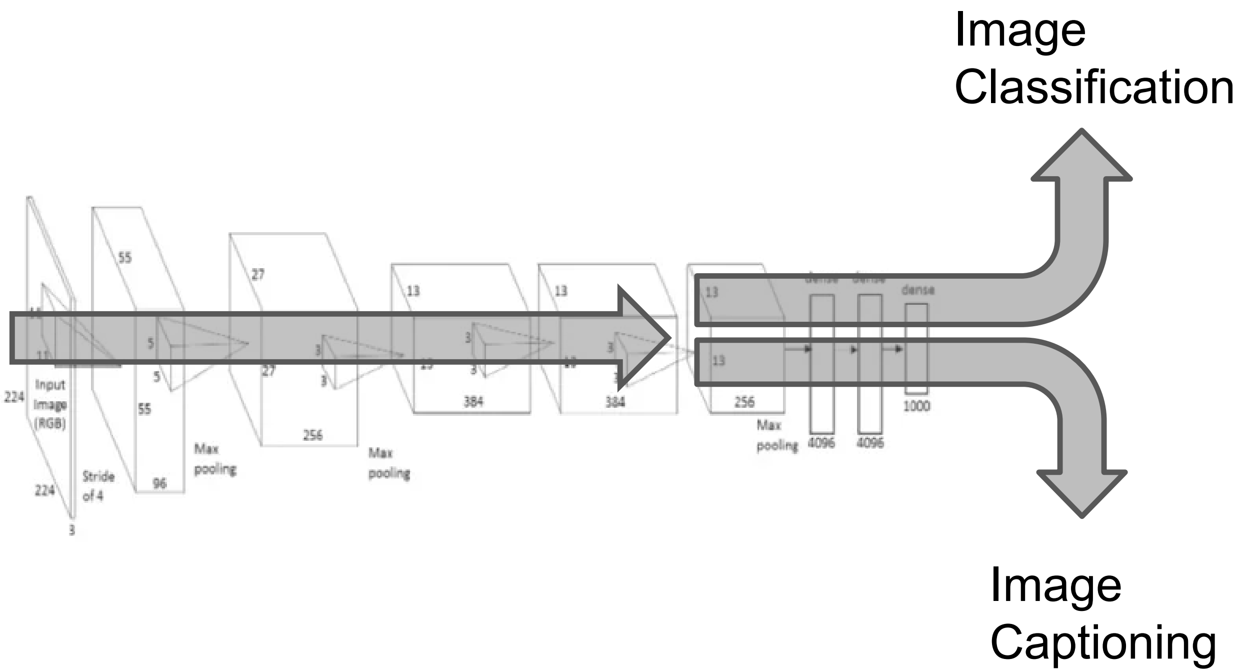


Freeze or slow weight training

Image Captioning

Multi-task Learning

Sharing model weight across tasks often improves performance on both.



Summary of Deep Net Properties

- Layered architecture (the deep part) of simple units.
- Inner layer representations are learned only from end-to-end tasks.
- Depth and complexity seem to be only limited by the amount of data. More complex models → better representations → better accuracy.
- This behavior is fundamentally different from classical ML: there is often no obvious performance ceiling.
- Inner layer representations are typically task-independent → easy to re-use models for applications that don't have large training datasets.
- Multi-task learning usually works: another departure from typical behavior of classical ML methods.