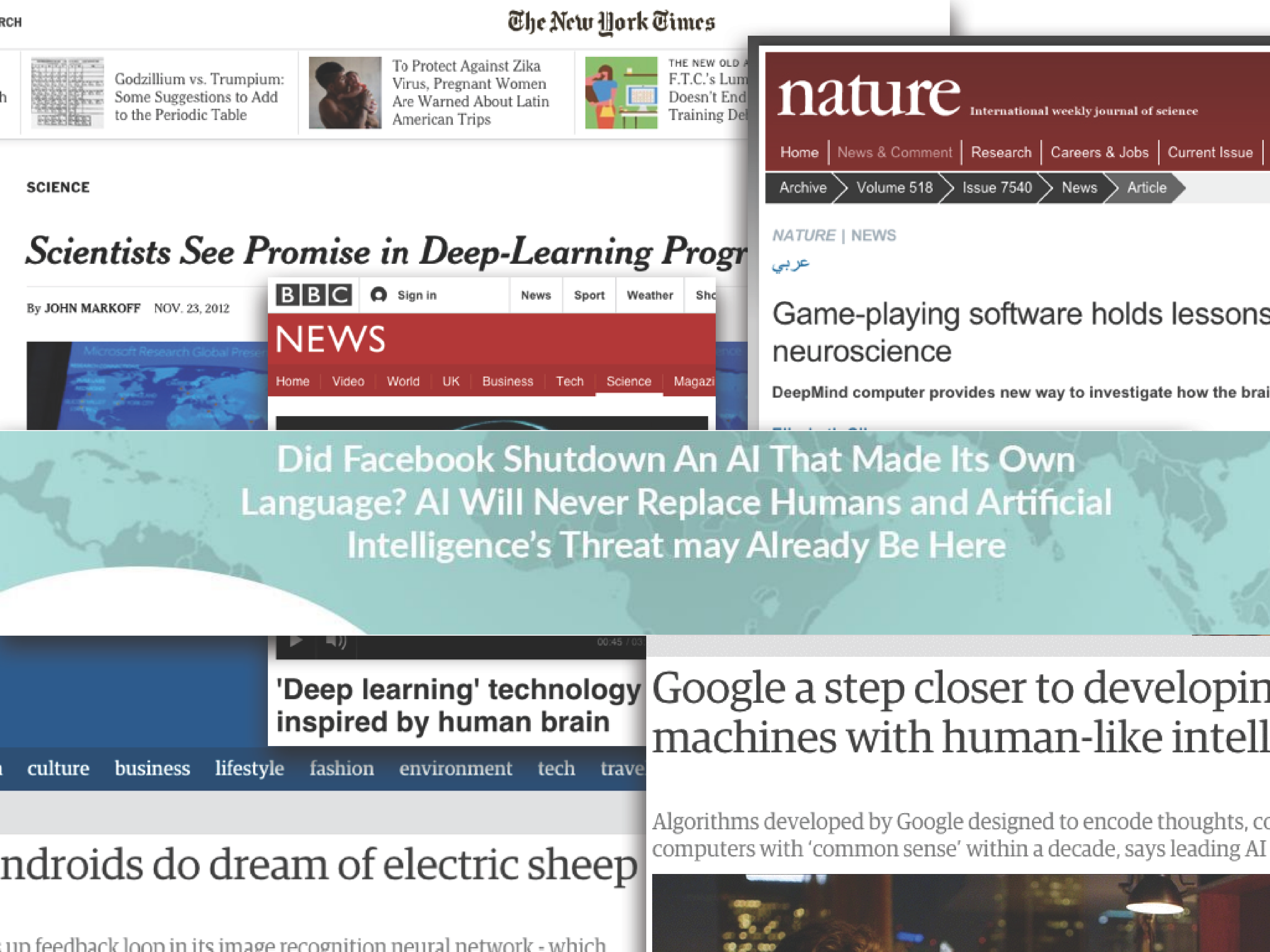
CS182/282A: Designing, Visualizing and Understanding Deep Neural Networks

John Canny

Spring 2019

CN: 182: 32191, 282A: 31116



The New York Times
Godzillium vs. Trumpium: Some Suggestions to Add to the Periodic Table
To Protect Against Zika Virus, Pregnant Women Are Warned About Latin American Trips
SCIENCE
Scientists See Promise in Deep-Learning Progr
By JOHN MARKOFF NOV. 23, 2012
nature
International weekly journal of science
Home
News & Comment
Research
Careers & Jobs
Current Issue
Archive
Volume 518
Issue 7540
News
Article
NATURE | NEWS
عربي
Game-playing software holds lessons
neuroscience
DeepMind computer provides new way to investigate how the brai
BBC
Sign in
News
Sport
Weather
NEWS
Home
Video
World
UK
Business
Tech
Science
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' technology inspired by human brain
culture
business
lifestyle
fashion
environment
tech
ndroids do dream of electric sheep
up feedback loop in its image recognition neural network - which
Google a step closer to developin
machines with human-like intell
Algorithms developed by Google designed to encode thoughts, co
computers with 'common sense' within a decade, says leading AI

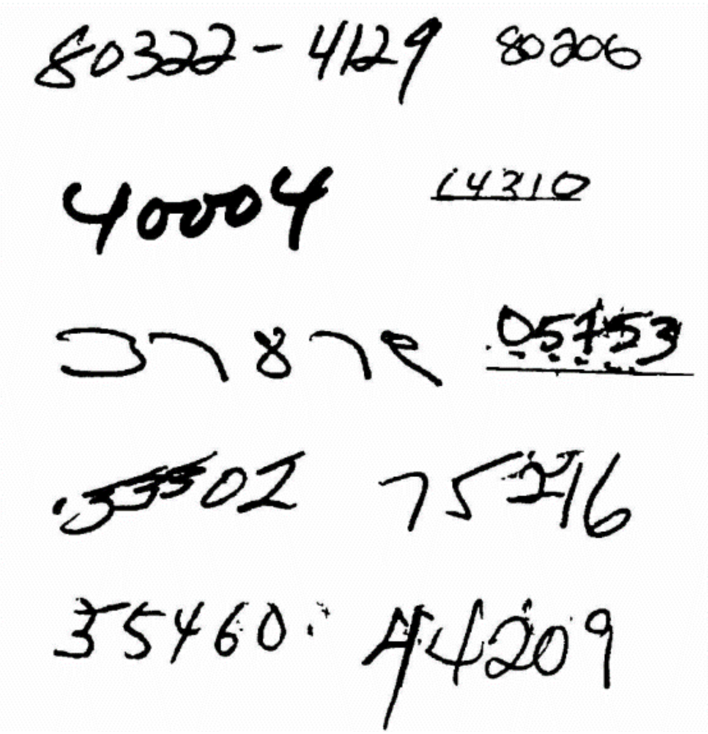
Deep Learning: Hype or Hope?

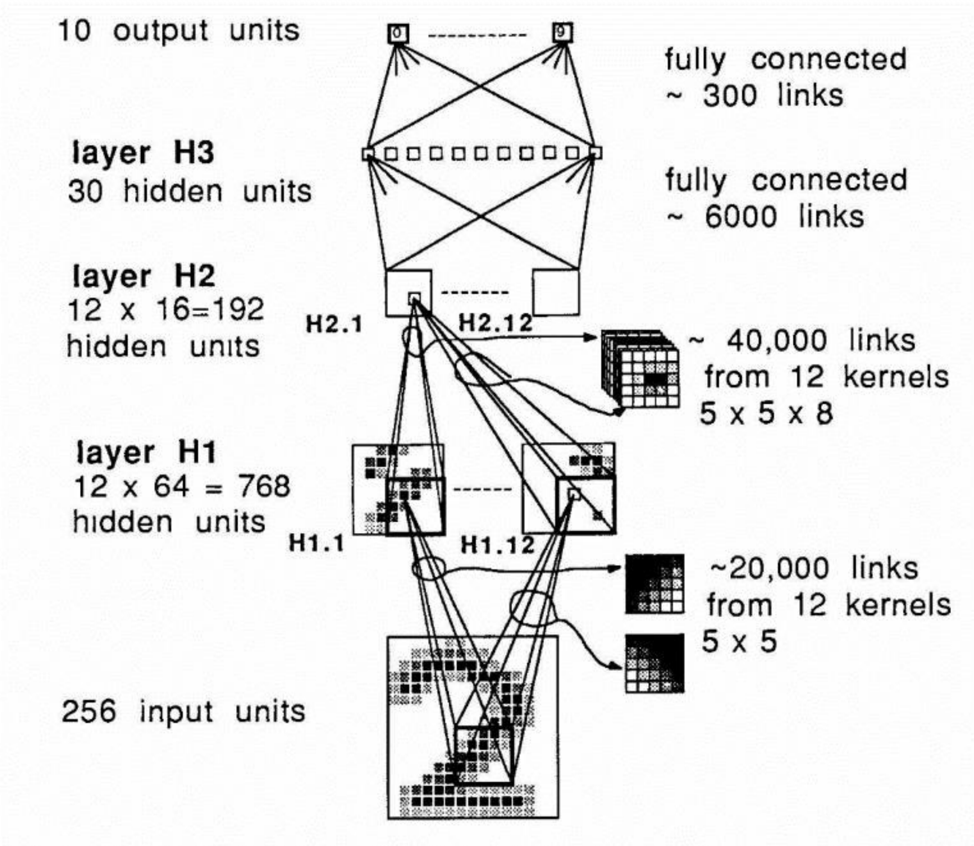
Hype: (n) “extravagant or intensive publicity or promotion”

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

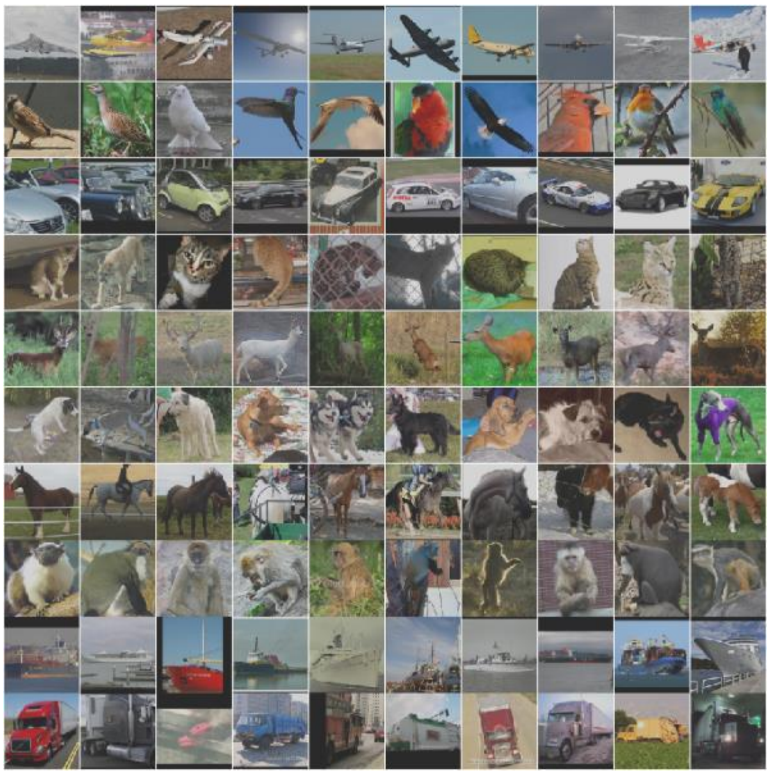


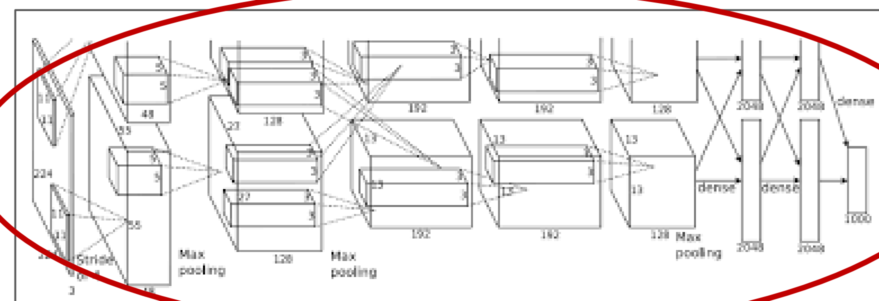


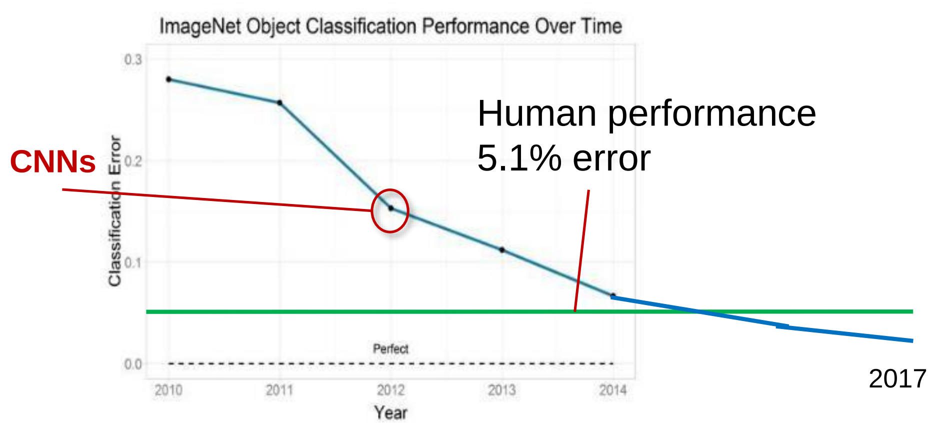
Milestones: Image Classification

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

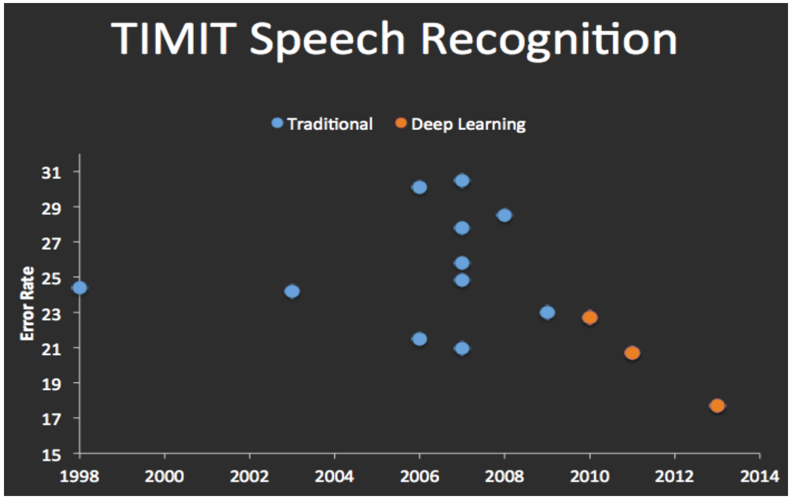
Almost a LeNet !

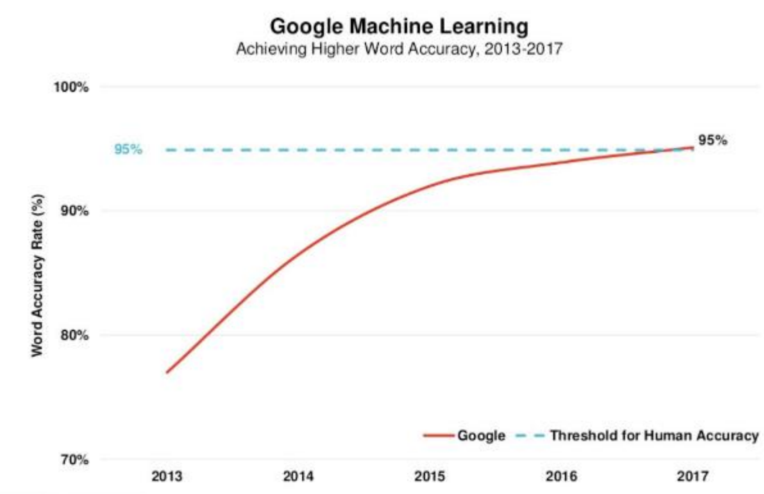






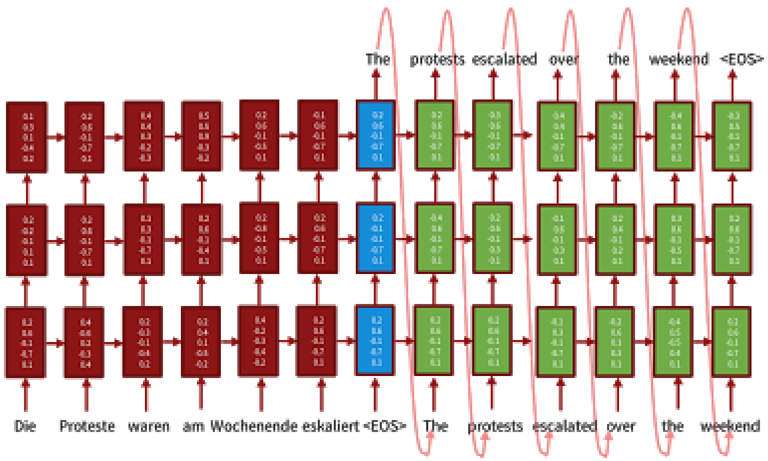
Milestones: Speech Recognition

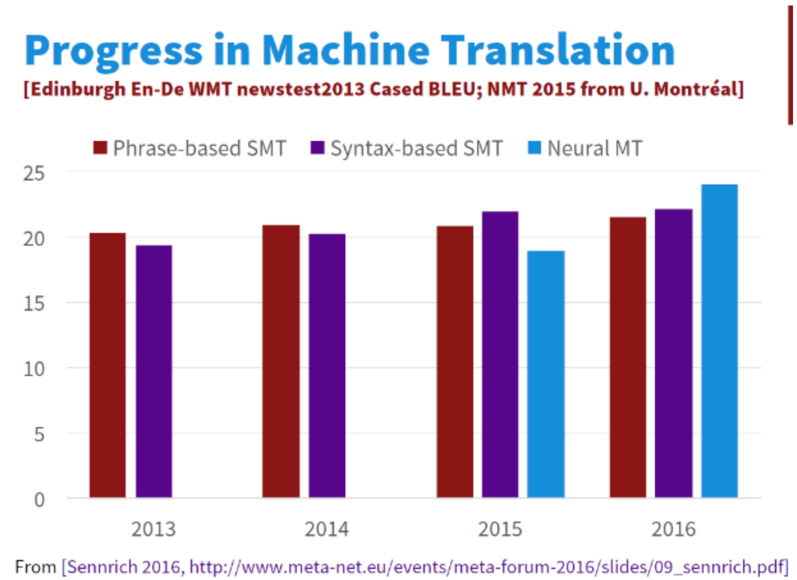




Milestones: Language Translation

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





Source Luong, Cho, Manning ACL Tutorial 2016.

Neural Text Processing 2017

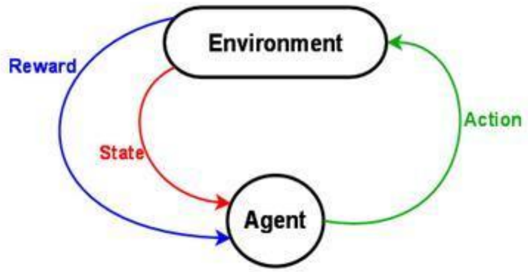
State of the art results on NLP application-level tasks

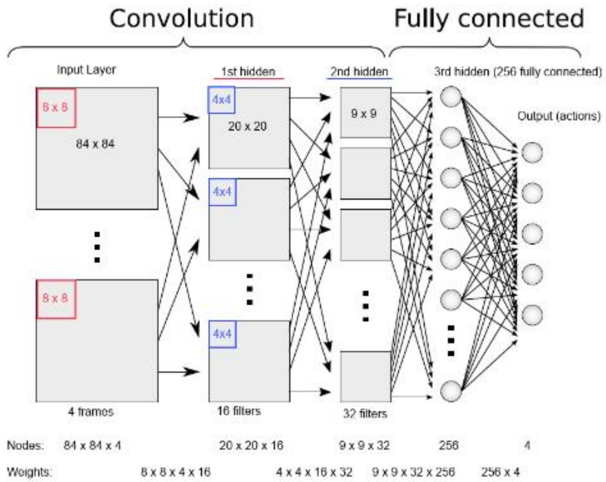
Task	Test set	Metric	Best non-neural	Best neural	Source
Machine Translation	Enu-deu newstest16	BLEU	31.4	34.8	http://matrix.statmt.org
	Deu-enu newstest16	BLEU	35.9	39.9	http://matrix.statmt.org
Sentiment Analysis	Stanford sentiment bank	5-class Accuracy	71.0	80.7	Socher+ 13
Question Answering	WebQuestions 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.

Risk 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?

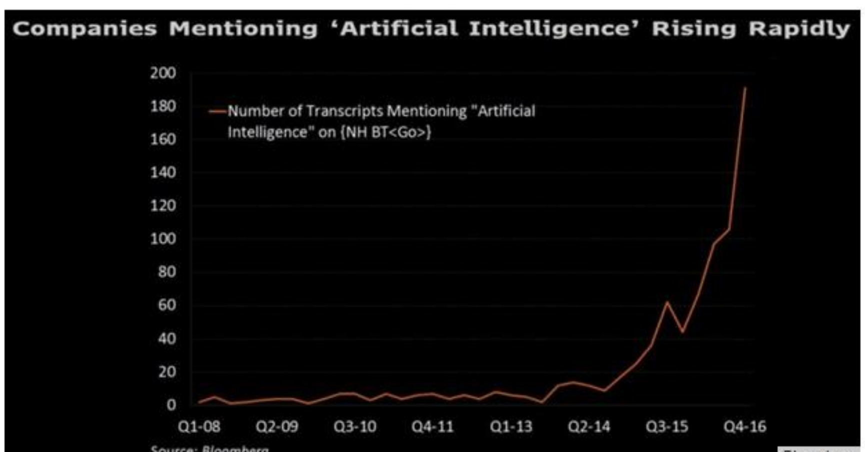
Forbes Technology Council

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

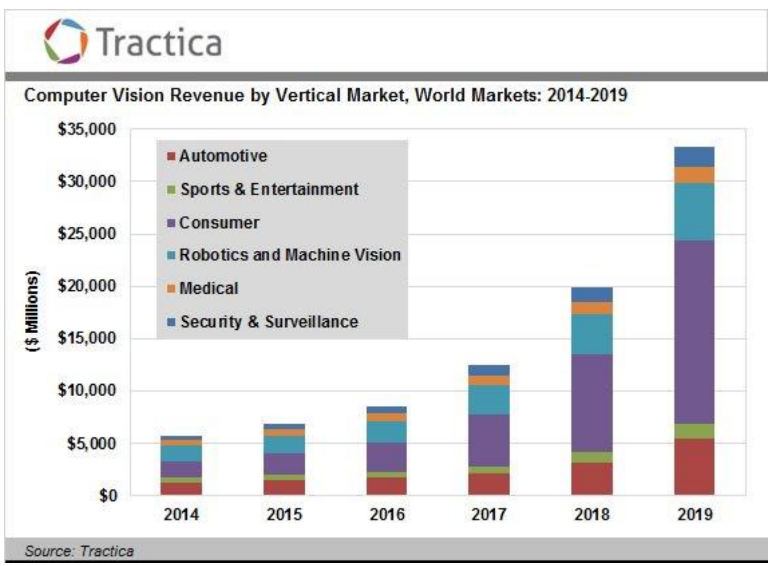
By Matt Asay, InfoWorld | MAR 3, 2017

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

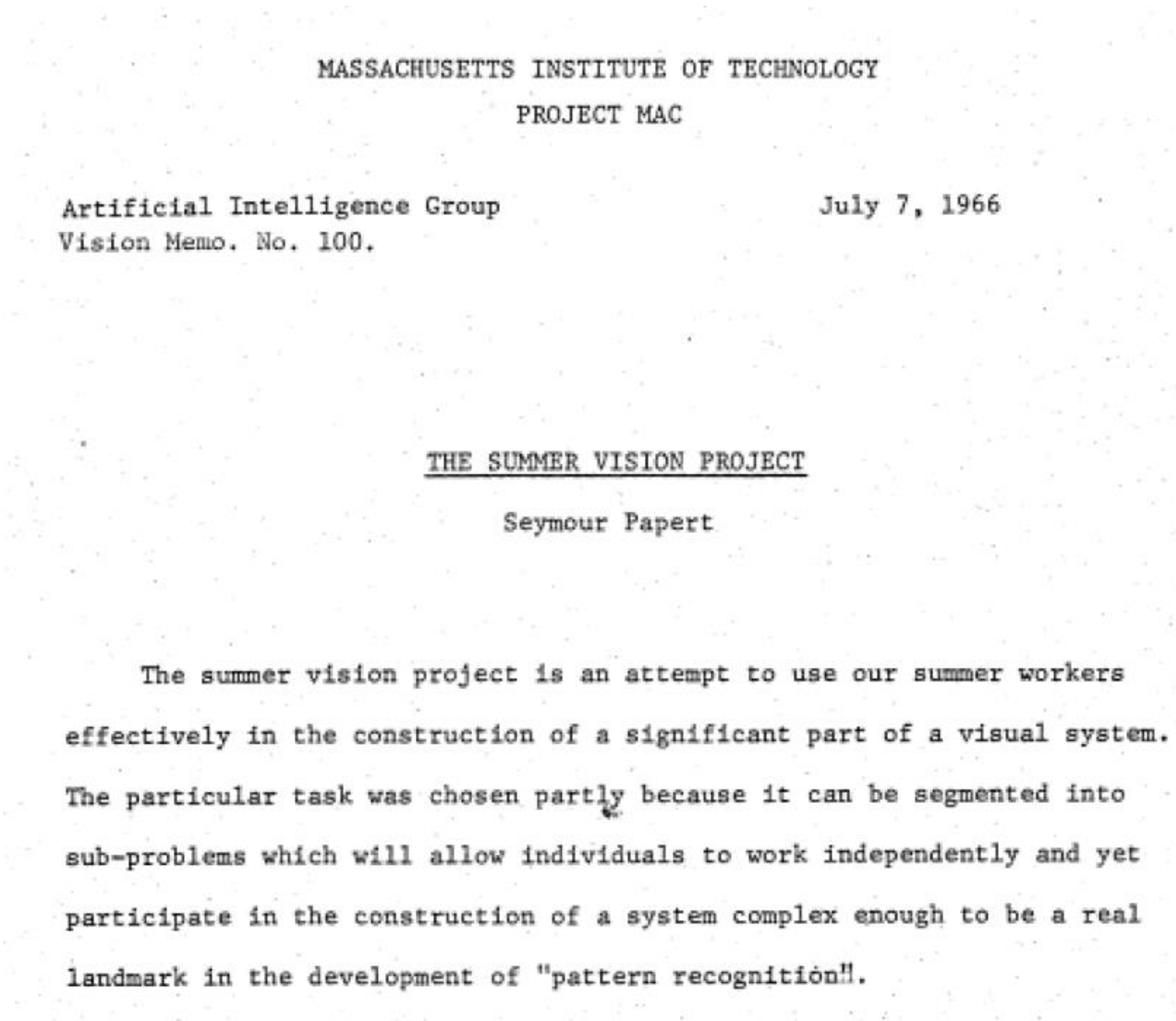
Elite CIOs, CTOs & execs offer firsthand insights on tech & business. FULL BIO

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.

THE SUMMER VISION PROJECT

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

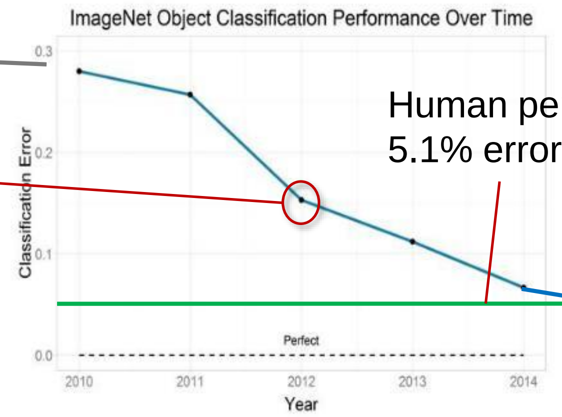
AI systems are still not “generally intelligent,” … even if they are awfully good at an awful lot of things.

- 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:

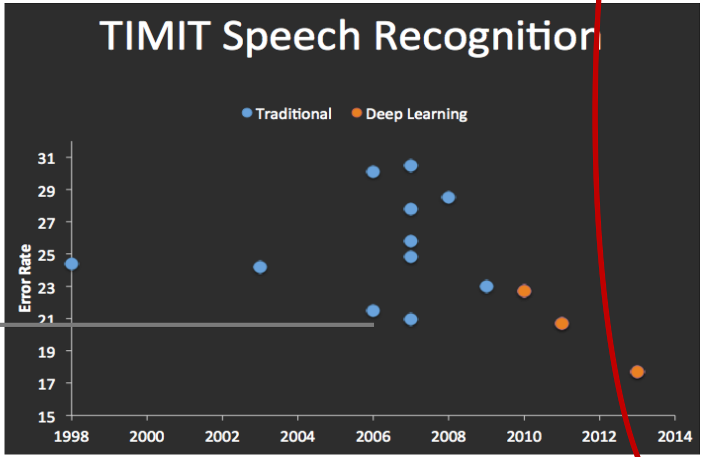
Performance floor

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)

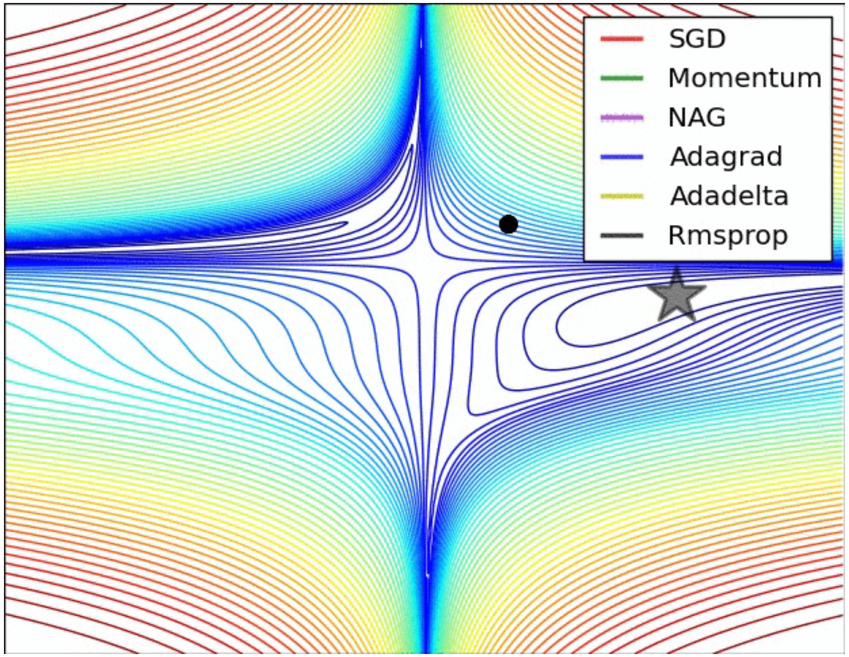
Goals:

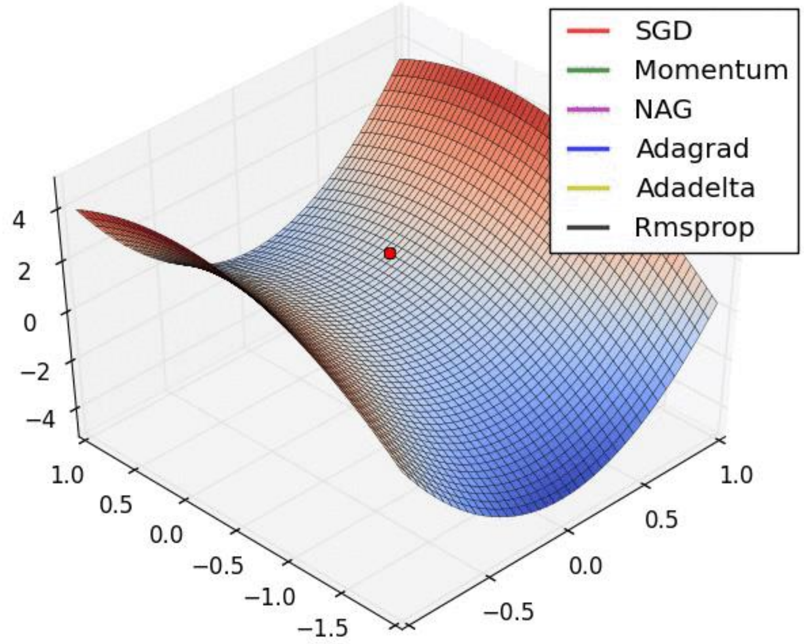
- Introduce deep learning to a broad audience.
- Review principles, techniques and visualization for understanding deep networks.
- Develop skill at designing networks for applications.

Materials:

- Book(s)
- Notes
- Lectures

The role of Animation

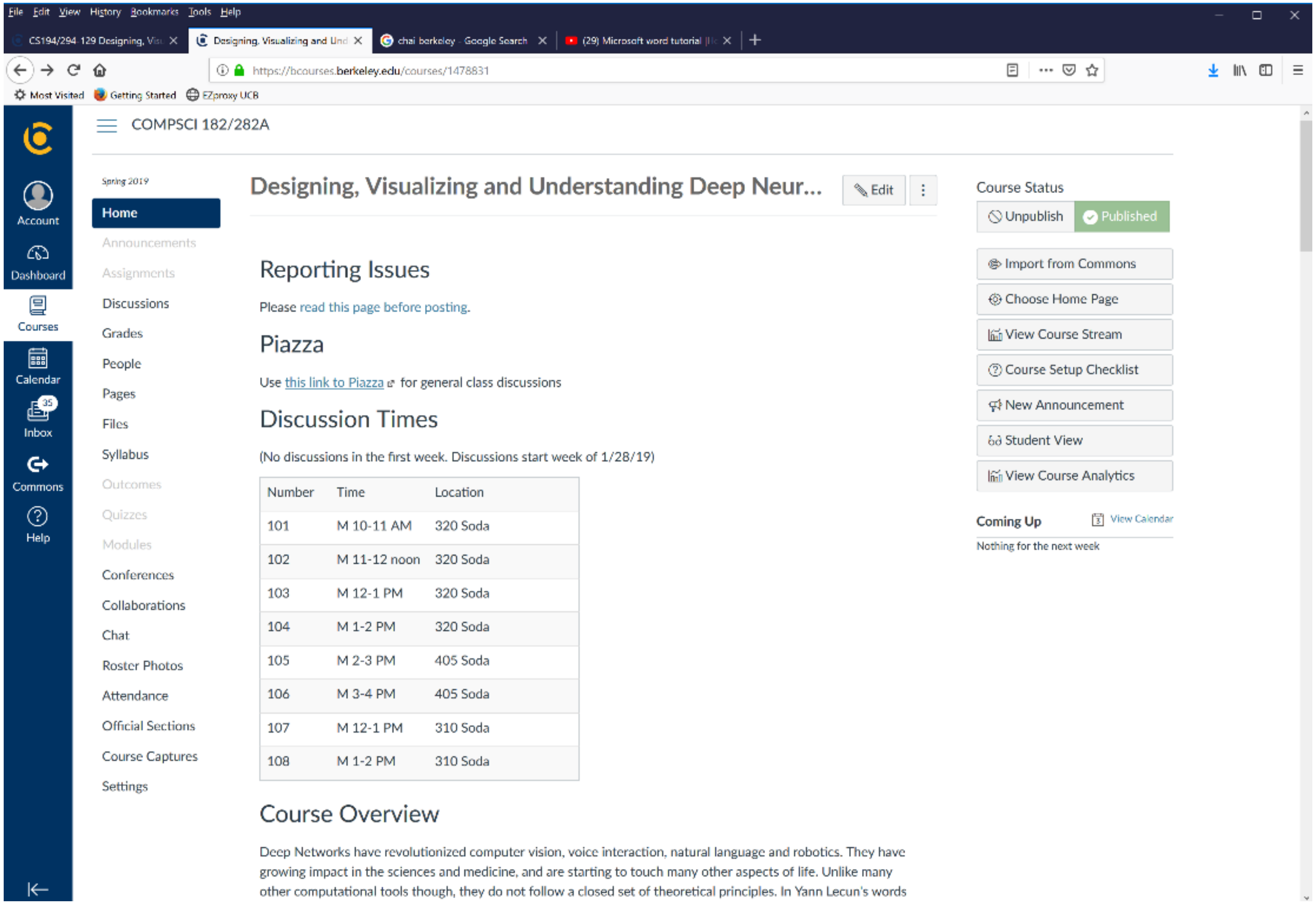




From A. Karpathy’s cs231n notes.

This Course: bCourses page

Please sign in to the bCourses, course info is on the main page.



bCourses main page has the schedule

ols Help

Designing, Visualizing and Und × | chai berkeley - Google Search × | (29) How To Convert pdf to wo ×

https://bcourses.berkeley.edu/courses/1478831

EZproxy UCB

The class will be screen-captured and audio-recorded. Recordings will be available from bCourses and CalCentral. You should be able to get them from the "Course Captures" tab at the left.

Schedule

Date	Lecture Topic	Reading	Assignments/Section Notes
W 1/23	Introduction, Course Overview, Brief history of deep networks.	Introduction from Deep Learning	
M 1/28	Machine learning concepts: loss and risk, discriminative models, linear and logistic regression.	Review: sections 1.1-1.3 and 6.6-6.9 from the CS189 book (skip KL-div). Do Python/Numpy tutorial if needed.	
W 1/30	SVMs, multiclass classification, softmax. Cross-validation.	sections 6.10-6.11, 1.6 from the CS189 book (skip Tikhonov)	Assignment 1 out
M 2/4	Optimization, Stochastic Gradient Descent. Robbins-Monro.	Chapter 8 of Deep Learning Optimization Notes Backpropagation Notes	
W 2/6	Backpropagation, Convolutional Networks.	Convnet notes	
M 2/11	CNN examples, activation functions, initialization, batch normalization.	Convnet notes Training Neural Networks 1 Training Neural Networks 2	Project Proposal out
W 2/13	Training: Batch normalization, dropout, ensembles, hyperparameter tuning.	Training Neural Networks 3	

bCourses Schedule (home page)

Intro., ML review

Computer Vision, General DNN principles

Natural language, Generative and Adversarial Nets

Imitation and reinforcement learning

Date	Lecture Topic	Reading	Assignments/Section Notes
W 1/23	Introduction, Course Overview, Brief history of deep networks.	Introduction from Deep Learning	
M 1/28	Machine learning concepts: loss and risk, discriminative models, linear and logistic regression.	Review: sections 1.1-1.3 and 6.6-6.9 from the CS189 book (skip KL-div). Do Python/Numpy tutorial if needed	
W 1/30	SVMs, multiclass classification, softmax. Cross-validation.	sections 6.10-6.11, 1.6 from the CS189 book (skip Tikhonov)	Assignment 1 out
M 2/4	Optimization, Stochastic Gradient Descent, Robbins-Monro.	Chapter 8 of Deep Learning Optimization Notes Backpropagation Notes	
W 2/6	Backpropagation, Convolutional Networks.	Convnet notes	
M 2/11	CNN examples, activation functions, initialization, batch normalization.	Convnet notes Training Neural Networks 1 Training Neural Networks 2	Project Proposal out
W 2/13	Training: Batch normalization, dropout, ensembles, hyperparameter tuning.	Training Neural Networks 3	
M 2/18	Holiday: President's Day		
Tu 2/19			Assignment 1 due 11pm Assignment 2 out
W 2/20	Object Detection and Segmentation.		Practice Midterm 1
M 2/25	Recurrent Networks, LSTMs, applications.	RNN chapter from Deep Learning Understanding LSTM Networks	Project proposal due
W 2/27	Visualizing Deep Networks.	Quite a few visualizations will be covered. Browse this list	
M 3/4	Midterm 1		
W 3/6	Semantic Models for Text.	WordToVec Skip-Thought Vectors	
M 3/11	Attention Networks	Recurrent Models of Visual Attention	Assignment 2 due 11pm Assignment 3 out
W 3/13	Natural Language Translation.	NMT by Jointly Learning to Align and Translate Attention is All You Need	
M 3/18	Memory Networks, Text Question-Answering Systems.	End-to-End Memory Networks Learning End-to-End Goal Directed Dialog	
W 3/20	Relational and Graph modeling networks		
3/25-29	Spring Break		
M 4/1	Adversarial and Fooling Networks.	Adversarial Examples Transferability/blackbox attacks Physical perturbation	Practice Midterm 2
W 4/3	Generative Models: Autoregressive and VAEs.		Assignment 3 due Assignment 4 out
M 4/8	Generative Adversarial Networks.	Generative Adversarial Networks	
W 4/10	Midterm 2		
M 4/15	Imitation Learning.	End-to-End Learning for Self-Driving Cars DAGGER Learning Transferable Policies From Virtual Demonstration to Real-World Manipulation	
W 4/17	Reinforcement Learning: Policy Gradients.	Policy Gradient Methods, chapter 13 of "Reinforcement Learning"	
M 4/22	Reinforcement Learning: Value-based methods.	DQN paper (Nature) Asynchronous Methods for Deep Reinforcement Learning	
W 4/24	Imagination and Curiosity.	Unifying Count-Based Exploration and Intrinsic Motivation Curiosity-driven Exploration by Self-supervised Prediction Imagination-Augmented Agents for Deep Reinforcement Learning	Assignment 4 due 11pm
M 4/29	Learning to Learn.	Learning to Learn blog post Matching Networks for One-Shot Learning Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks	
W 5/1	Playing Games		

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 Monday.

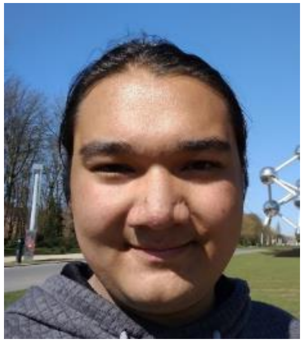
Course Staff

Prof: John Canny

GSIs:

Philippe Laban

Daniel Seita



David Chan

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 182/282A, SP 2019
- On bCourses, most material is publicly readable.
- Course numbers 182: 32191, 282A: 31116
- Instructor: John Canny lastname@berkeley.edu
- Time: MW 8am - 9:30am
- Location: 145 Dwinelle
- Discussion: Join Piazza for announcements and to ask questions about the course
- Section times, staff office hours are on the main page
- Webcasts (slides + audio) will be available from the bCourses page on left menu “Course Captures”

Course Project

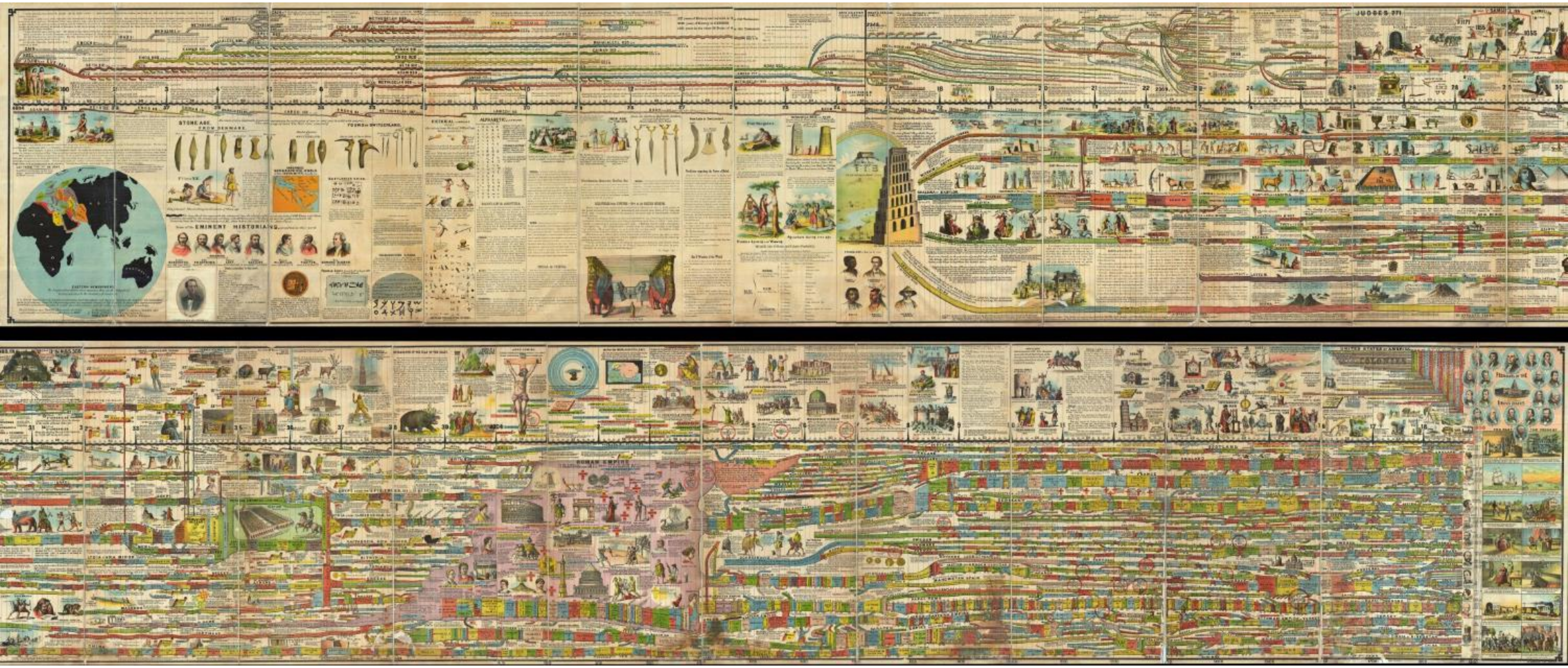
- Will consume about 2/3 of the semester.
- In teams of 3-4.
- Can be combined with other course projects
- We encourage “open-source” projects that can be archived.
- You will “check-in” with the GSIs several times during the semester.
- Final poster, video and report due at the end of the semester.

Questions ?

Coming up: Some rationale for deep neural networks…

Some History

- Reading: the Deep Learning Book, Introduction

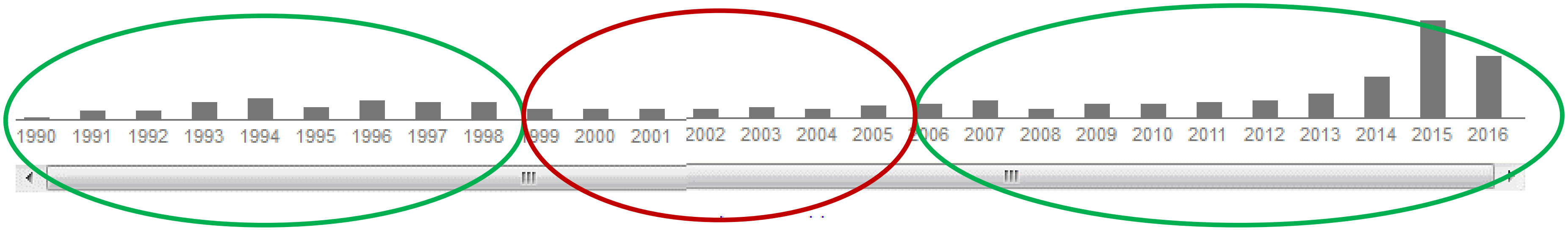


Phases of Neural Network Research

- 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:

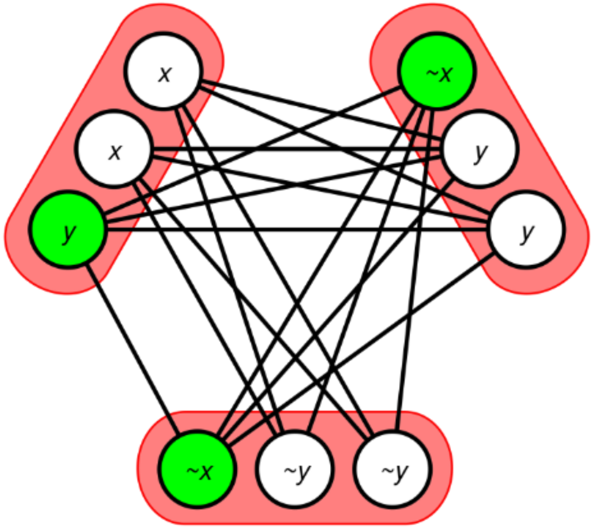


Second phase Deep Learning "Winter" Third phase

- 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.

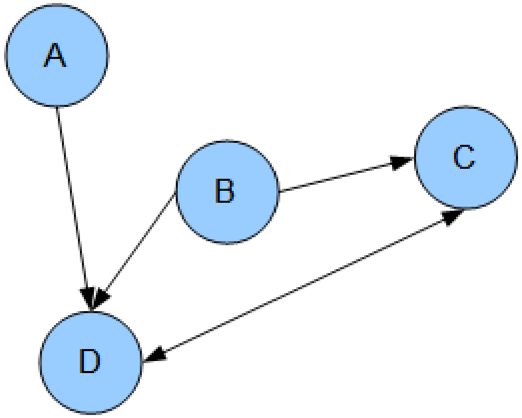
- Greedy algorithms are even more limited in what they can represent and how well they learn.

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

Why the success of DNNs is surprising

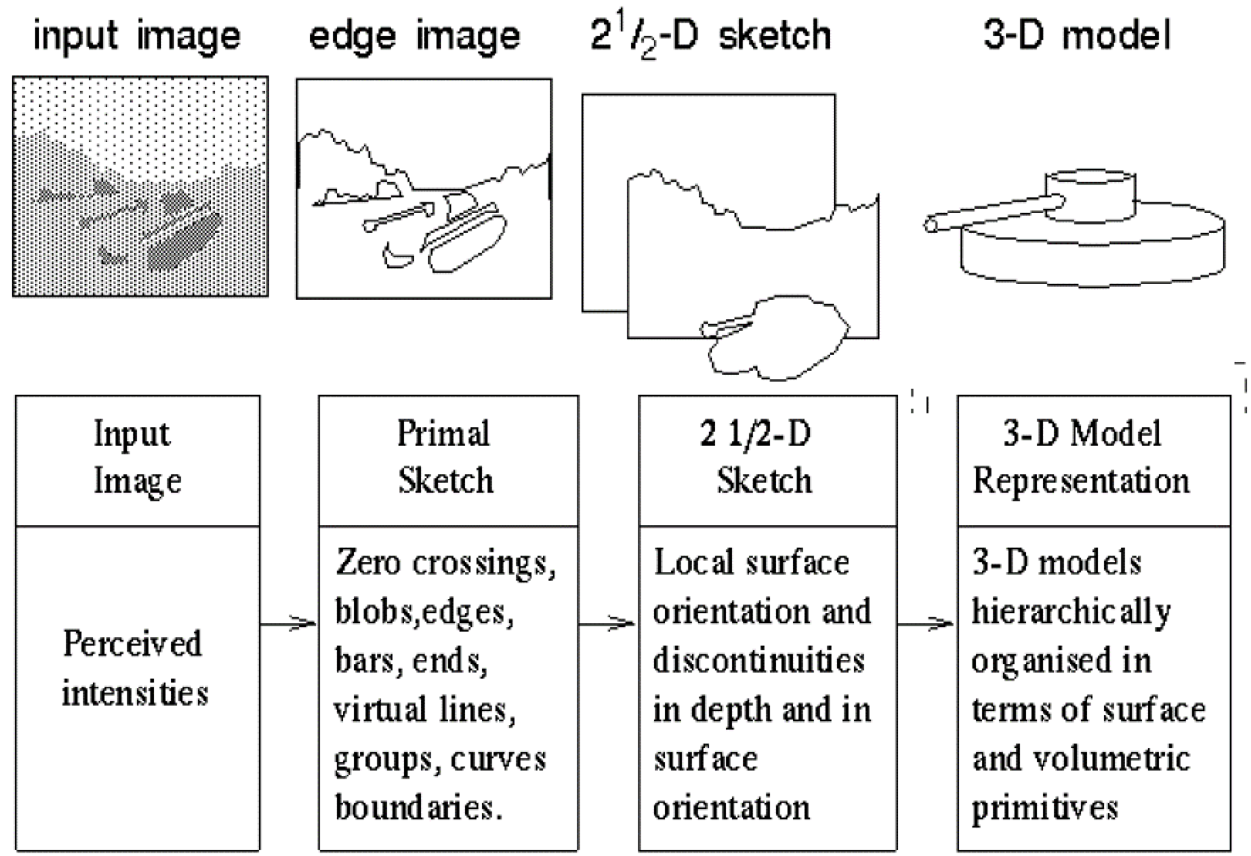
- In graphical models, values in a network represent random variables, and have a clear meaning. The network structure encodes dependency information, i.e. you can represent rich models.

- 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

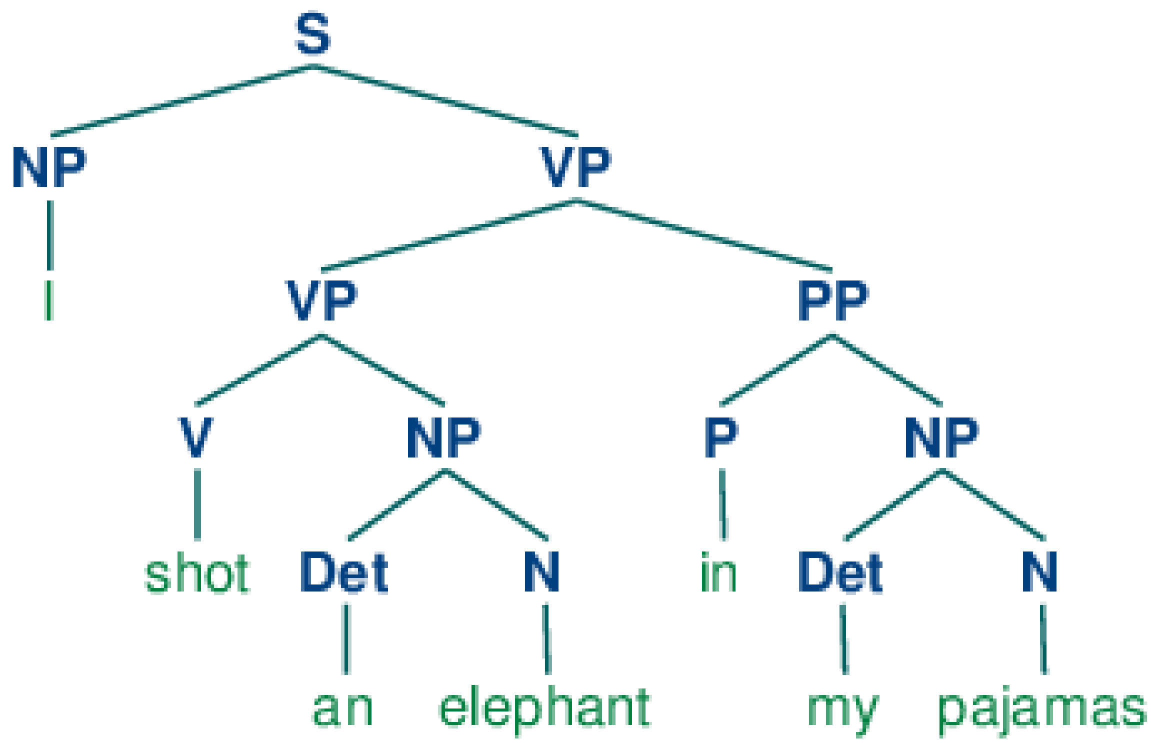
- Hierarchical representations are ubiquitous in AI. Computer vision:



Stages of Visual Representation, David Marr, 1970s

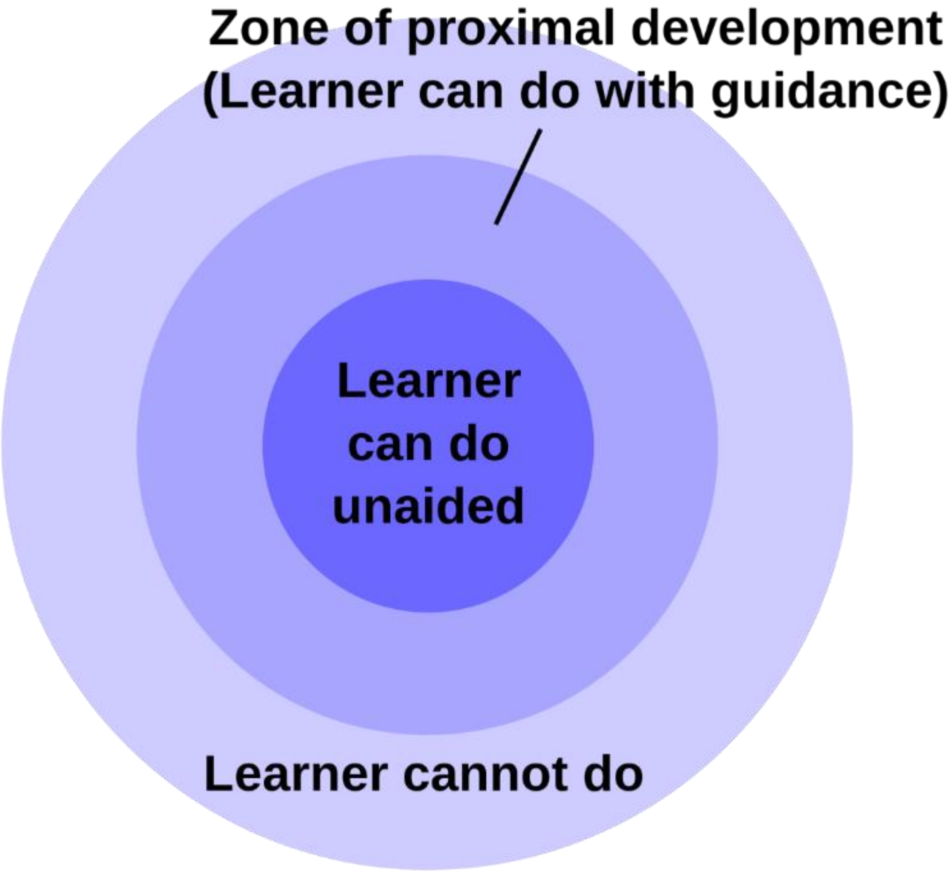
Why the success of DNNs is ~~surprising~~ obvious

- Hierarchies are ubiquitous in natural language:



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

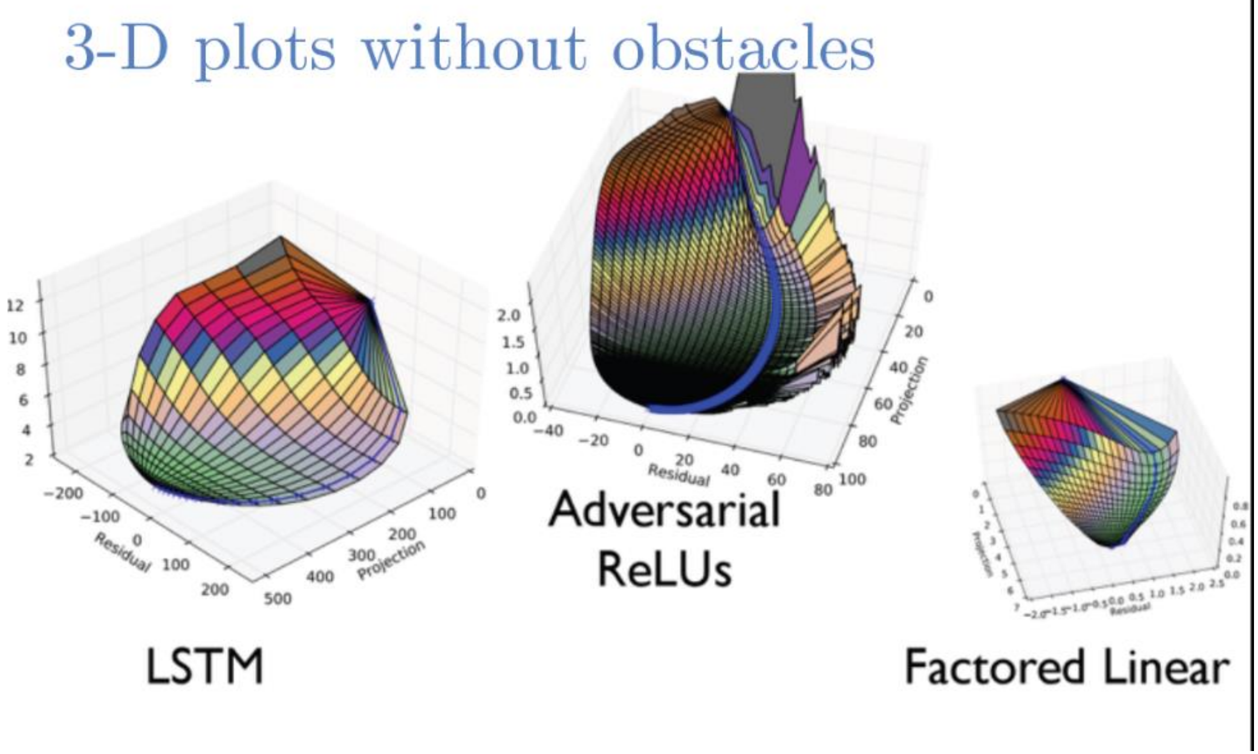
- Human Learning: is deeply layered.



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.



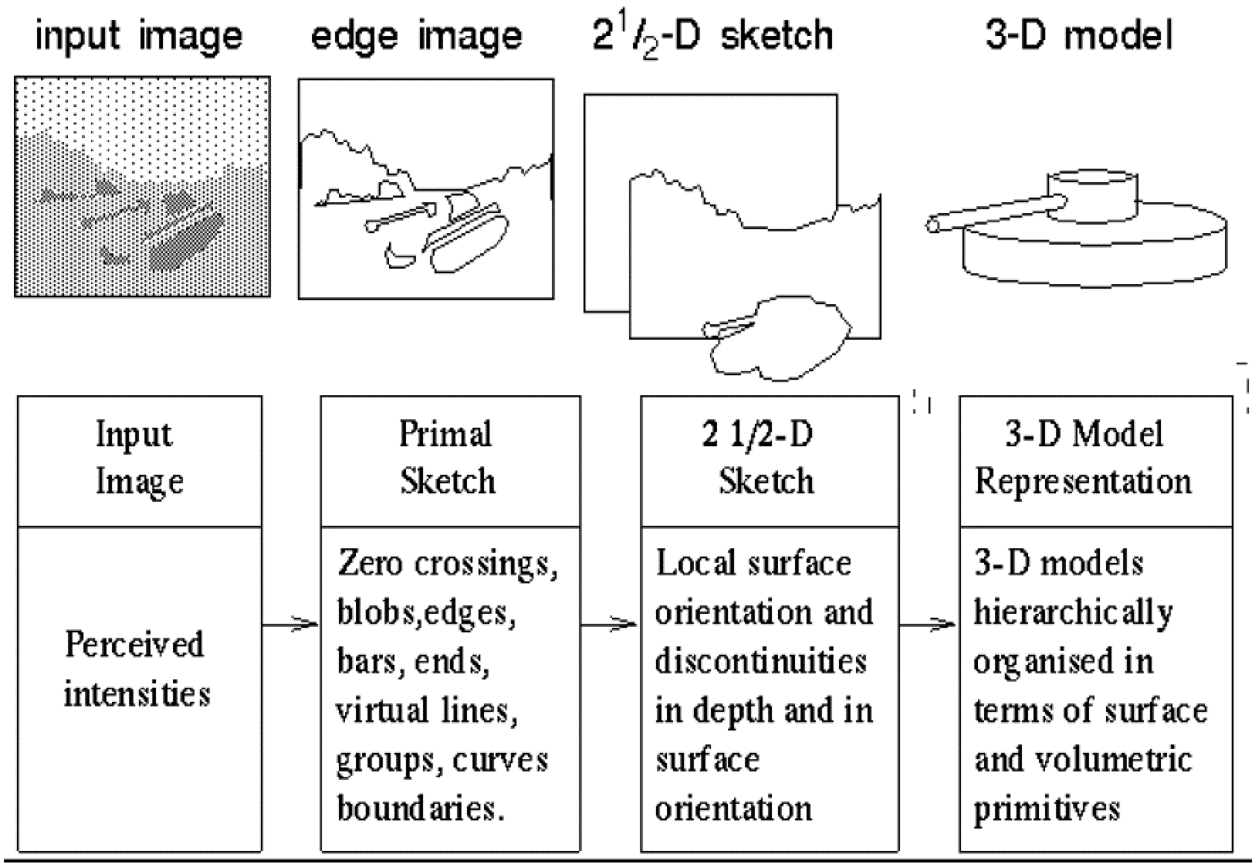
Ian Goodfellow – ICLR 2015 Tutorial

Questions ?

Coming up: Representation Learning

Hierarchical Representations

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

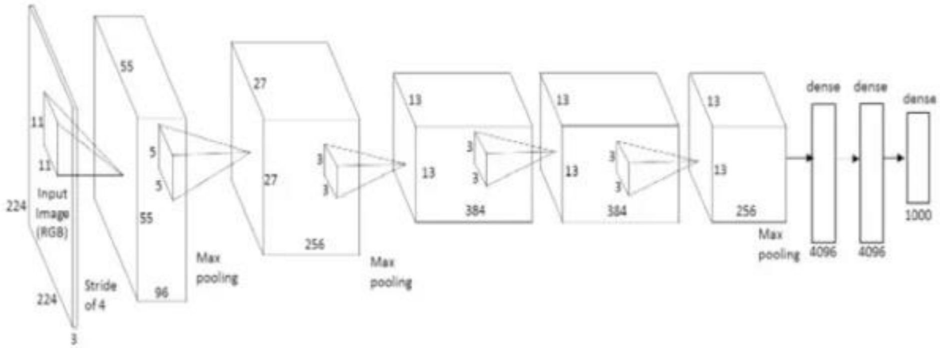


Stages of Visual Representation, David Marr, 1970s

Representation Learning

Learning Layered Representations: Training a deep Imagenet classifier:

1.4 million images



“Alexnet” (Krizhevsky) from 2012

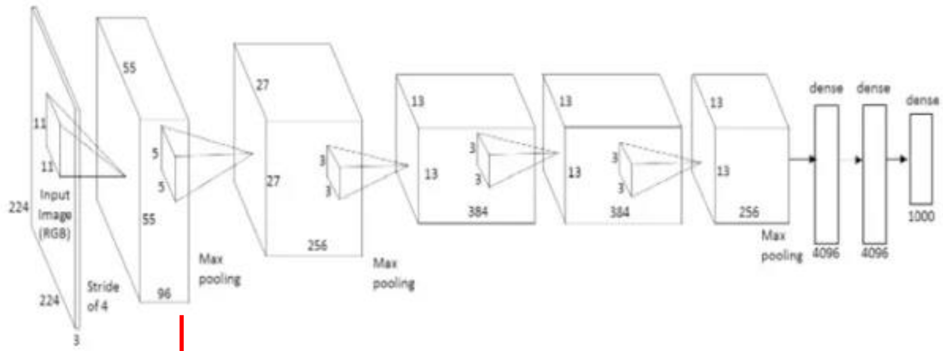
imagenet1000_clsid_to_human.txt

```
{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,
 6: 'stingray',
 7: 'cock',
 8: 'hen',
 9: 'ostrich, Struthio camelus',
 10: 'brambling, Fringilla montifringill
 11: 'goldfinch, Carduelis carduelis',
 12: 'house finch, linnet, Carpodacus me
 13: 'junco, snowbird',
 14: 'indigo bunting, indigo finch, indi
 15: 'robin, American robin, Turdus migr
 16: 'bulbul',
 17: 'jay',
 18: 'magpie',
 19: 'chickadee',
 20: 'water ouzel, dipper',
 21: 'kite',
 22: 'bald eagle, American eagle, Haliae
 23: 'vulture',
 24: 'great grey owl, great gray owl, St
```

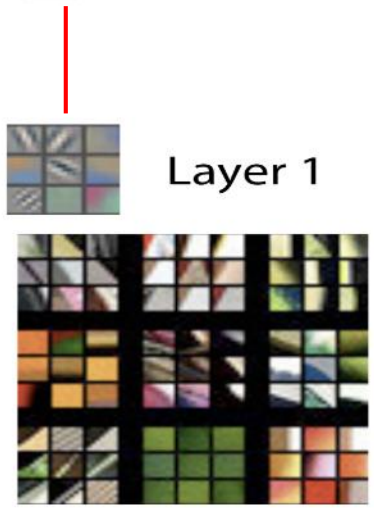
1000 labels

Representation Learning

After a lot of training…



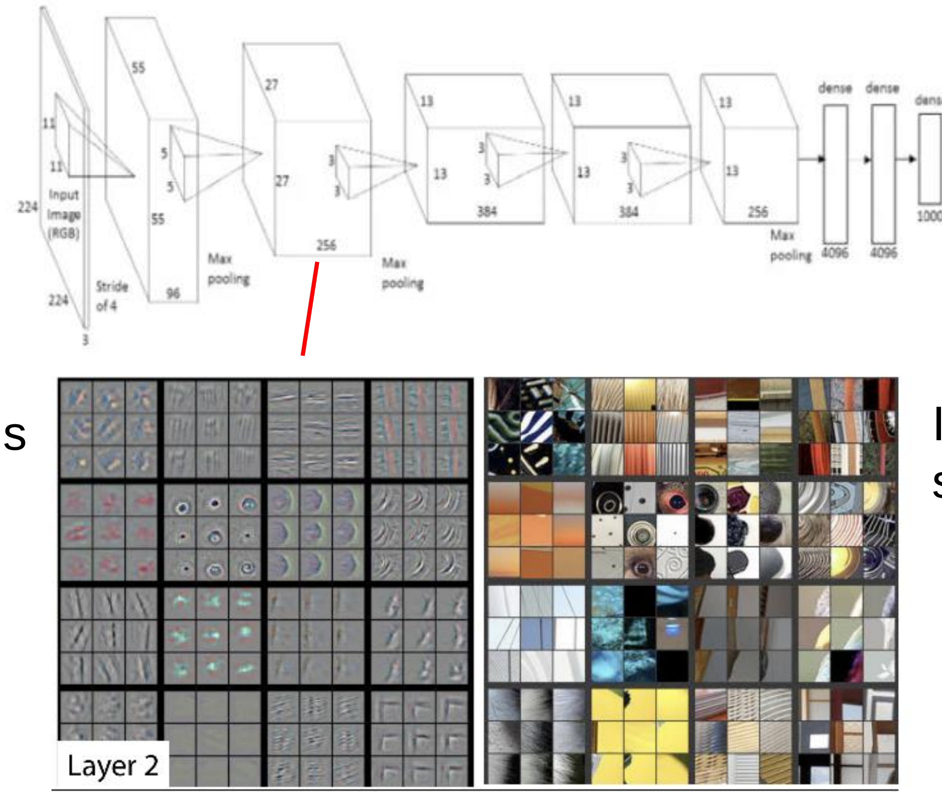
Neural receptive field
gray is zero
black is negative
white is positive



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

Image patches with strongest responses

Representation Learning

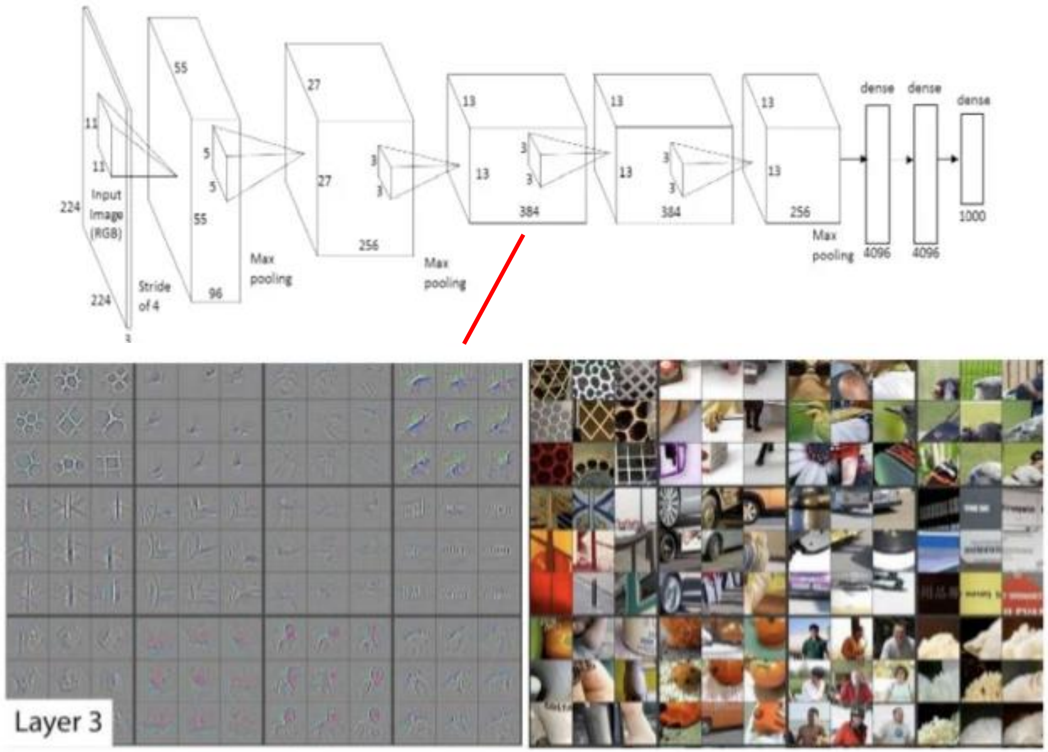


Neural receptive fields
gray is zero
black is negative
white is positive

Image patches with strongest responses

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

Representation Learning

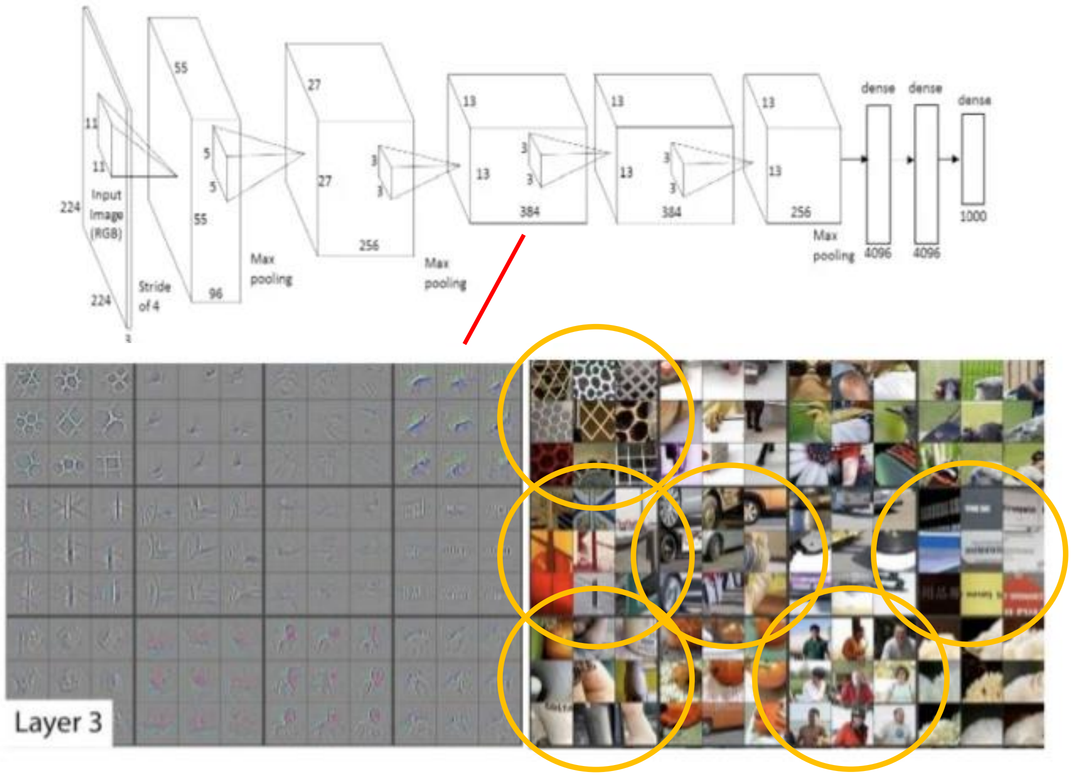


Neural receptive fields

Image patches with strongest responses

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

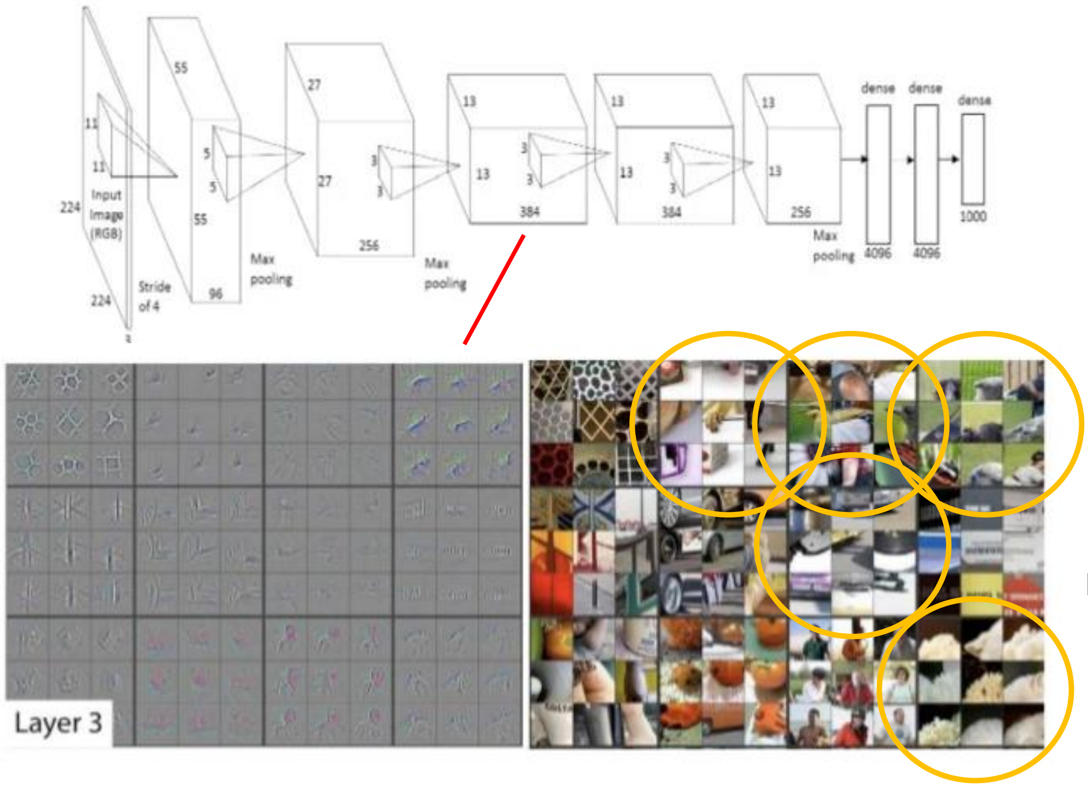
Representation Learning



Interpretable patches

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

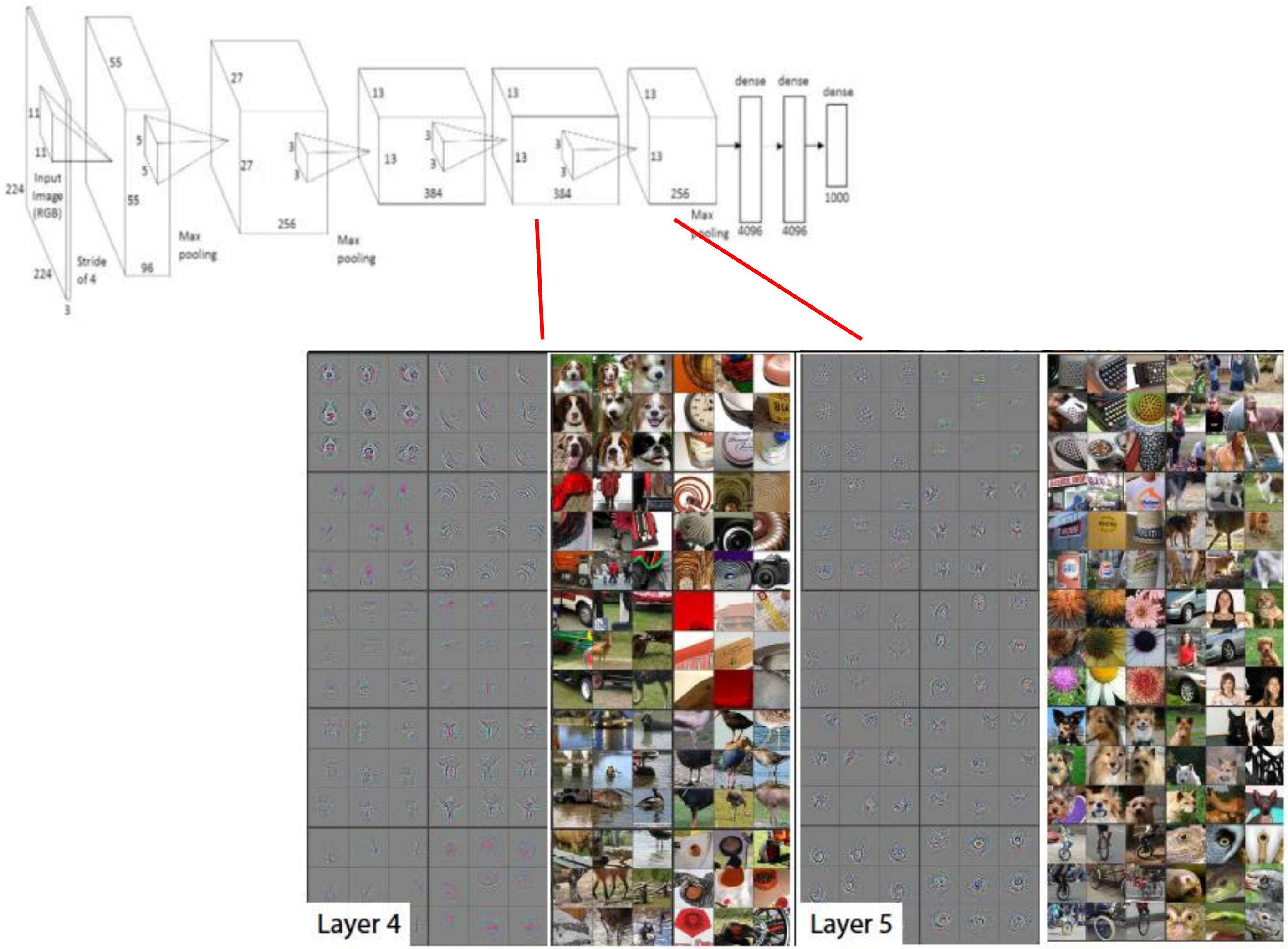
Representation Learning



Not so much…

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

Representation Learning



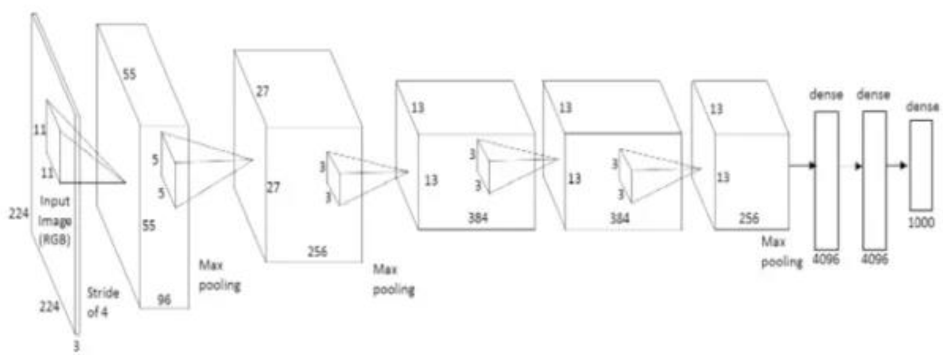
Zeiler and Fergus, “Visualizing and Understanding Convolutional Networks”, 2013

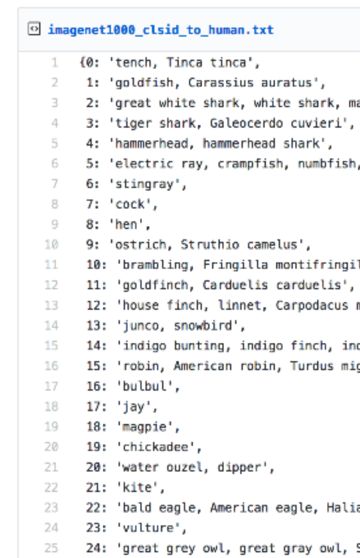
Representation Learning

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

Early and intermediate layer weights are typically task-independent.

1.4 million images





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

Consequences

You can reuse or customize models trained for different task.

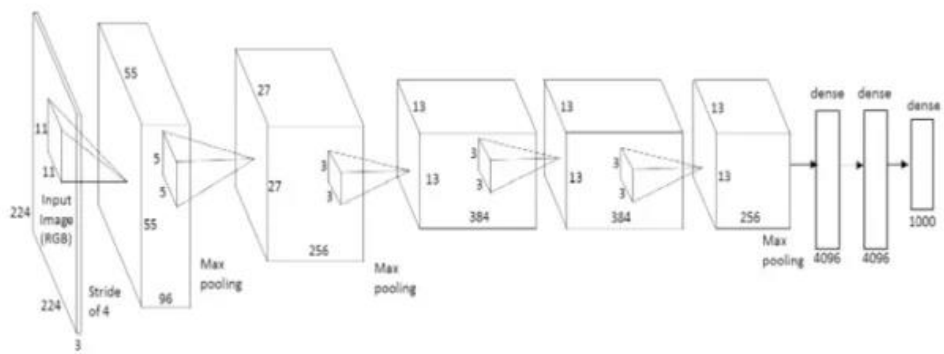
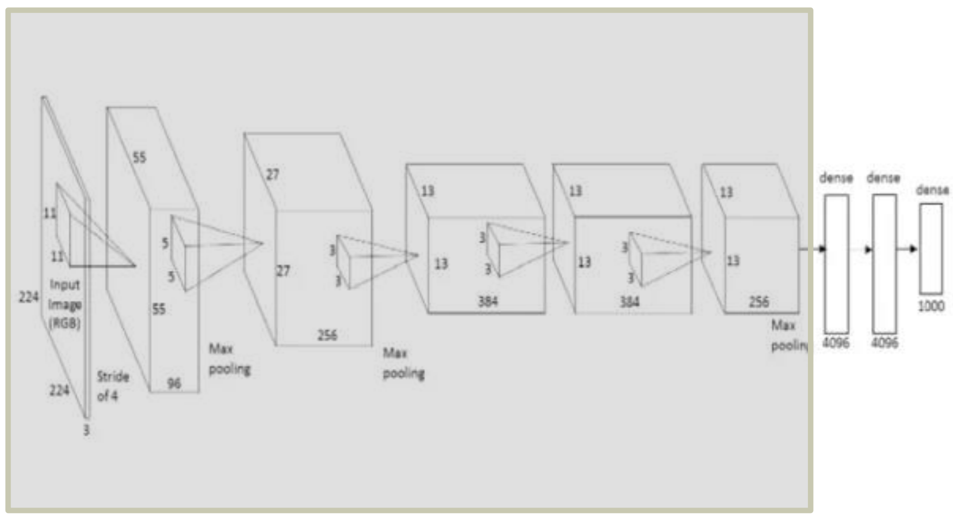


Image
Classification

Reuse/Fine Tuning

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

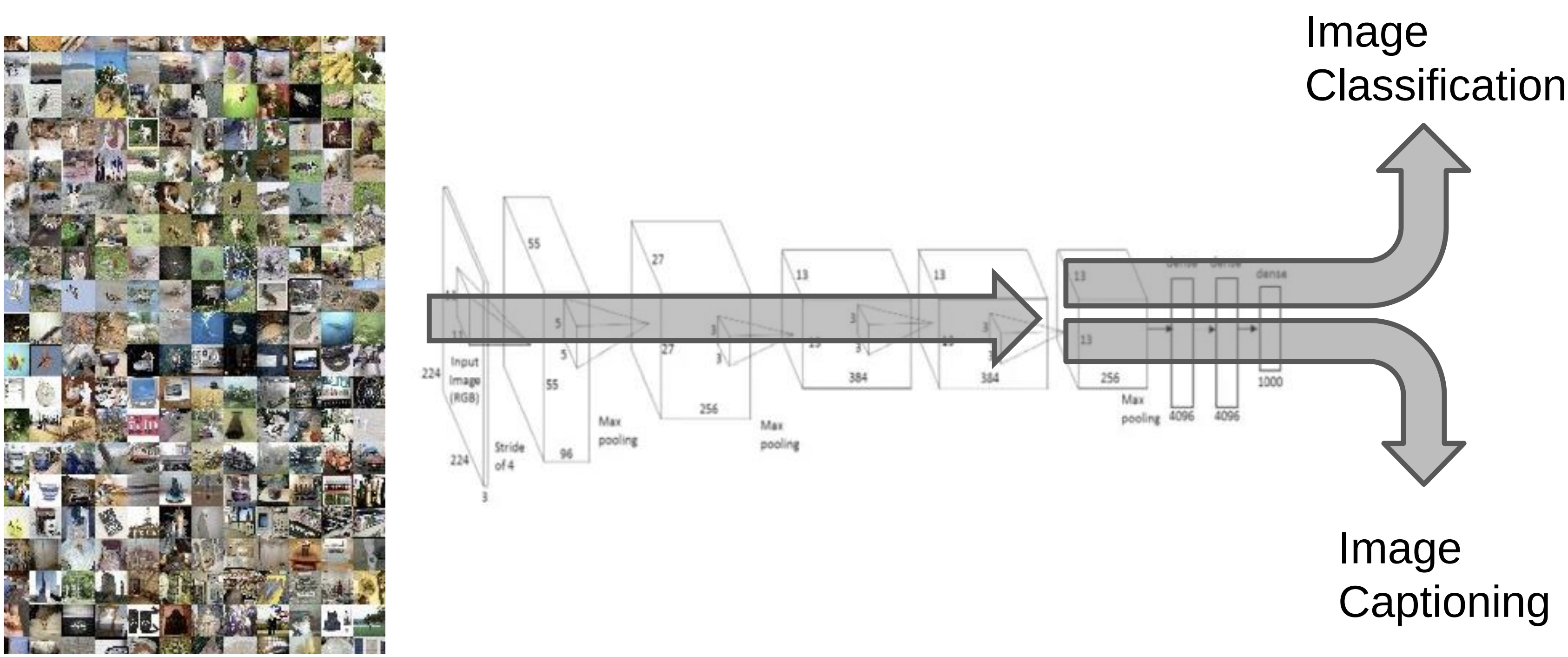


Freeze or slow weight training

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.