

14.332.435/16.332.530

Introduction to Deep Learning

Lecture 1 Introduction

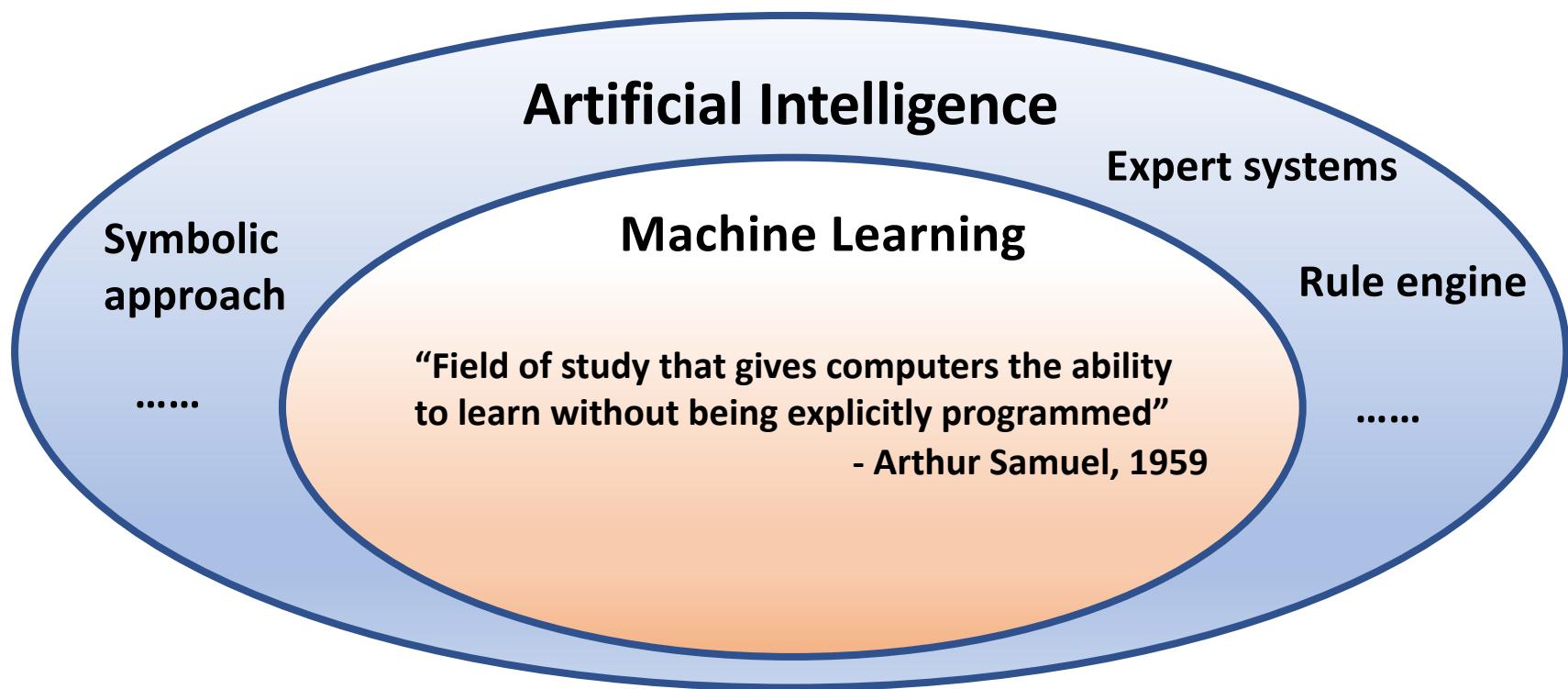
Yuqian Zhang

Department of Electrical and Computer Engineering

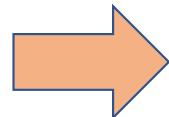
Today's Agenda

- Introduction to Deep Learning
- Course Information

What is Machine Learning

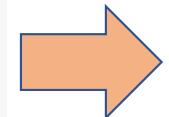


Example: Classification Task



An image classifier

```
def classify_image(image):  
    # Some magic here?  
    return class_label
```

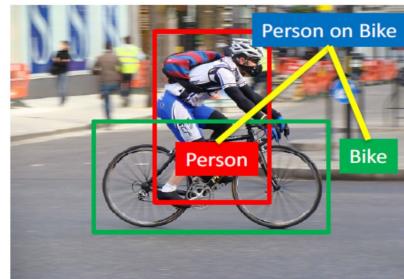
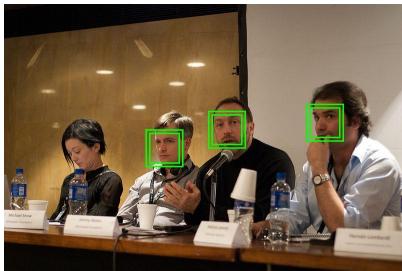


Plane

Unlike many tasks, e.g., sorting a list of numbers, no **obvious ways** to hard-code the algorithm for recognizing a plane or other classes.

Why Machine Learning

Such type of tasks is everywhere



Machine Learning: Data-driven Approach

```
def train(images, labels):
    # Machine Learning!
    return model
```

```
def predict(model, test_images):
    # Use model to predict labels
    return test_labels
```

Example training set

airplane



automobile



bird



cat



deer



1. Collect a dataset of images and labels
2. Use Machine Learning to train a classifier
3. Evaluate the classifier on new images

Machine Learning is Good At

- **Recognizing patterns**
 - Objects in real scenes
 - Facial identifies or facial expressions
 -
- **Recognizing anomalies**
 - Unusual sequences of credit card transactions
 - Unusual patterns of sensor readings in a nuclear power plant
 -
- **Prediction**
 - Future stock prices or currency exchange rates
 - Which movies will a person like?
 -
-

Q: What field is ML not good at?

History of Machine Learning

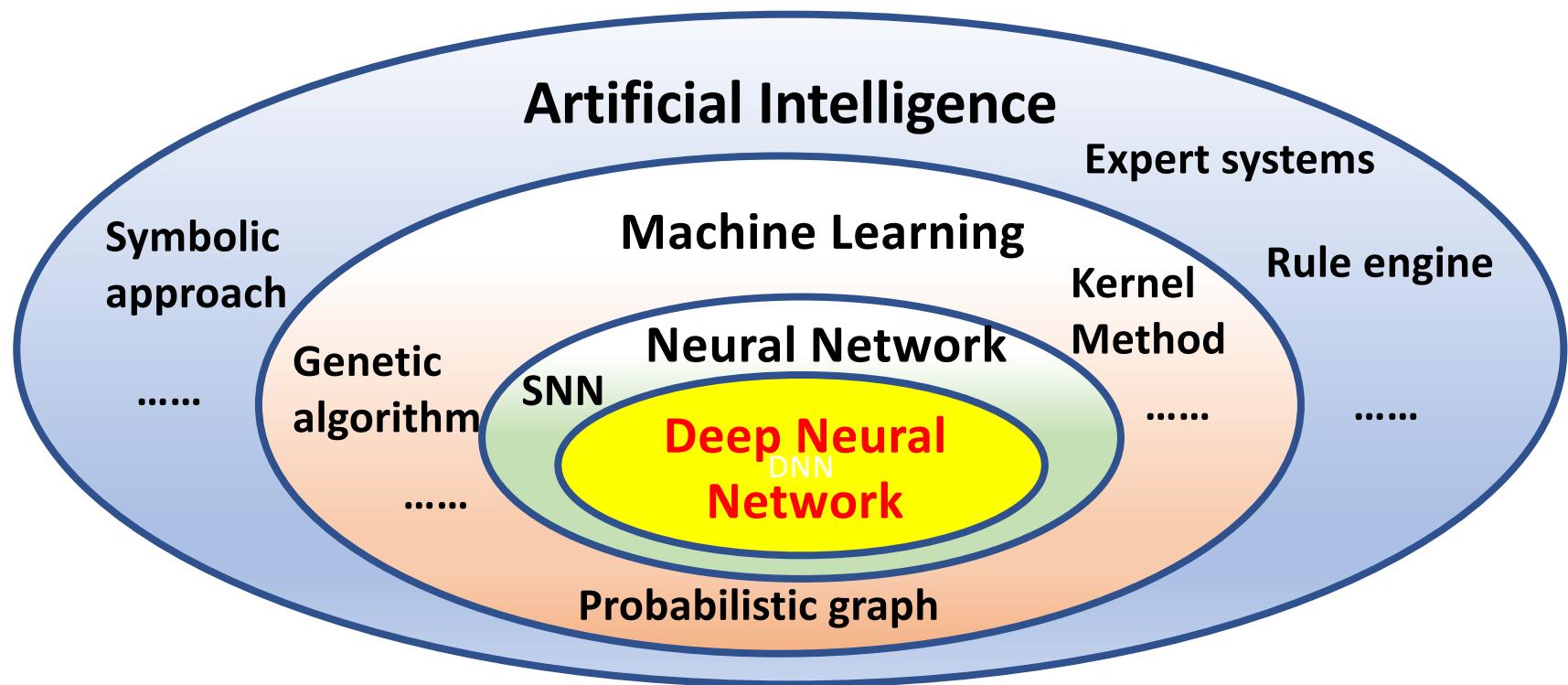
- 1957 — Perceptron algorithm (implemented as a circuit!)
- 1959 — Arthur Samuel wrote a learning-based checkers program that could defeat him
- 1969 — Minsky and Papert's book Perceptrons (limitations of linear models)
- 1980s — Some foundational ideas
- Connectionist psychologists explored neural models of cognition
 - 1984 — Leslie Valiant formalized the problem of learning as PAC learning
 - 1988 — Backpropagation (re-)discovered by Geoffrey Hinton and colleagues
 - 1988 — Judea Pearl's book Probabilistic Reasoning in Intelligent Systems introduced Bayesian networks

History of Machine Learning

- 1990s — the “AI Winter”, a time of pessimism and low funding
- But looking back, the ’90s were also sort of a golden age for ML research
 - Markov chain Monte Carlo
 - Variational inference
 - Kernels and support vector machines
 - Boosting
 - Convolutional networks
- 2000s — applied AI fields (vision, NLP, etc.) adopted ML
- 2010s — deep learning
 - 2010–2012 — neural nets smashed previous records in speech-to-text and object recognition
 - increasing adoption by the tech industry
 - 2016 — AlphaGo defeated the human Go champion
 - 2023 — ChatGPT

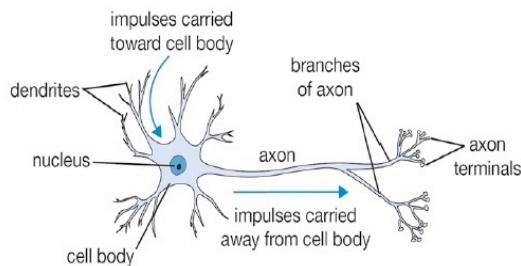
Focus of This Course: Deep learning

Deep Learning: (Deep) neural network

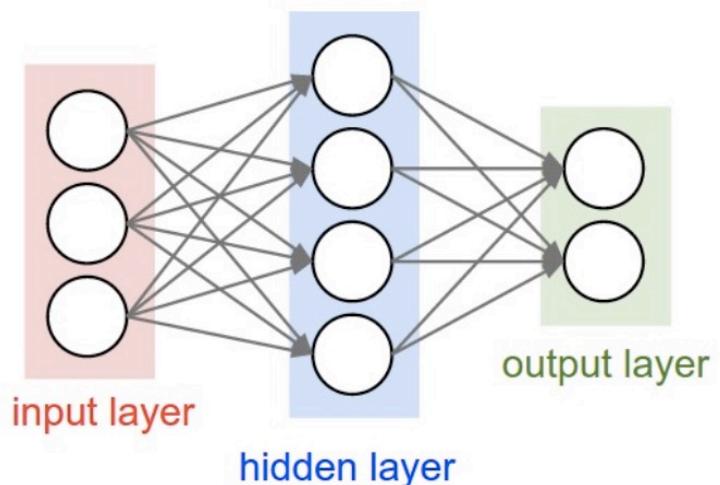
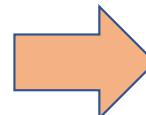


What is Deep Learning (Deep Neural Network)

- A type of artificial neural network
- “Deep Neural Network (DNN)”: more than 2 hidden layers.



Inspire



Artificial Neural Network

The basic computational unit of the brain is a **neuron**

→ 86B neurons in the brain

Neurons are connected with nearly $10^{14} - 10^{15}$ **synapses**

Neurons receive input signal from **dendrites** and produce output signal along **axon**, which interact with the dendrites of other neurons via **synaptic weights**

Synaptic weights – learnable & control influence strength

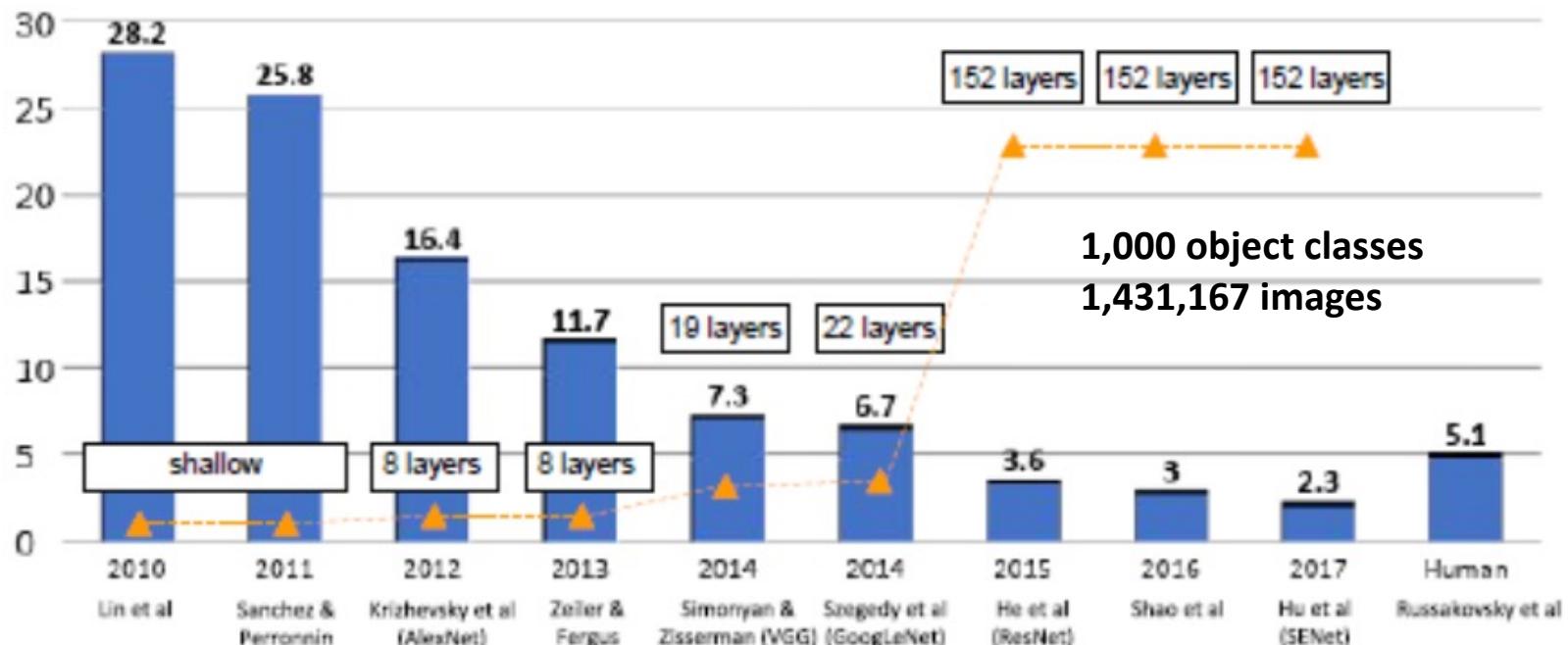
Why Deep learning

The most *dominated* ML algorithm

- Achieving record-breaking performance
- Huge social and media impact
- Widely popular and adopted in industry

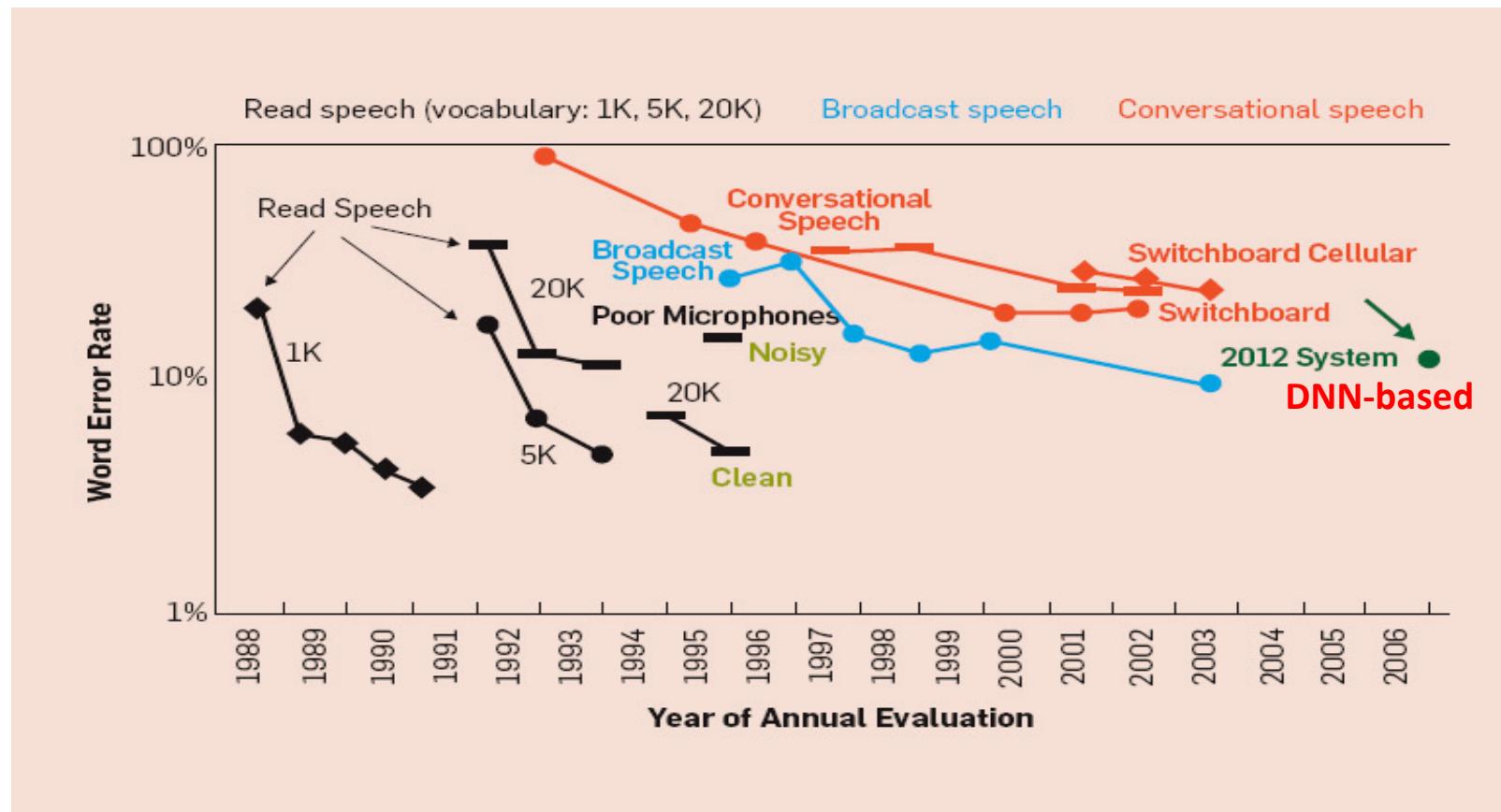
Record-breaking Performance (Vision)

ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners



Russakovsky et al. IJCV 2015

Record-breaking Performance (Speech)



Huang et al. CACM 2014

Record-breaking Performance (NLP)



Thang Luong
@lmthang

Follow

A new era of NLP has just begun a few days ago: large pretraining models (Transformer 24 layers, 1024 dim, 16 heads) + massive compute is all you need. BERT from [@GoogleAI](#): SOTA results on everything arxiv.org/abs/1810.04805. Results on SQuAD are just mind-blowing. Fun time ahead!

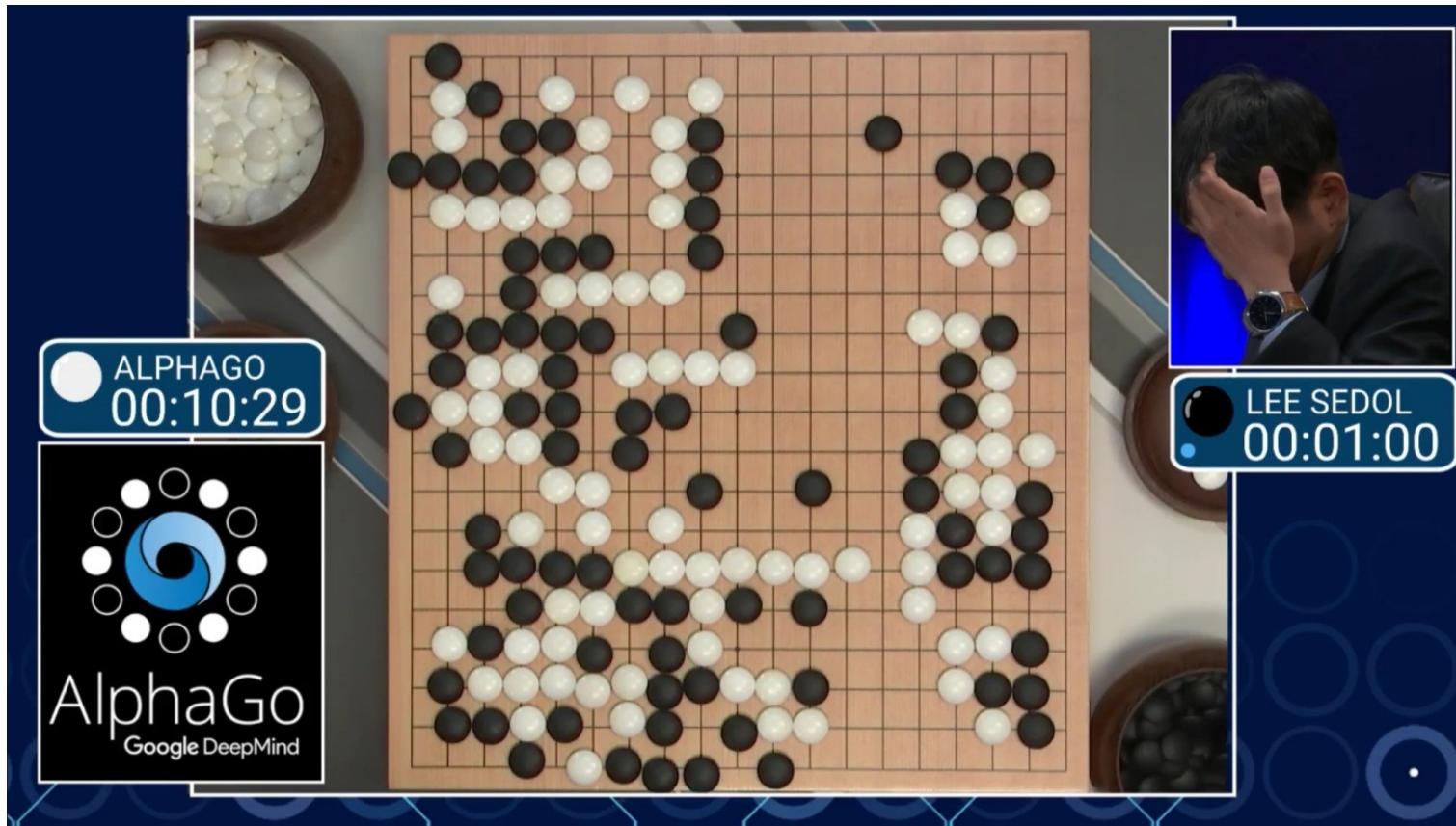
SQuAD1.1 Leaderboard

Since the release of SQuAD1.0, the community has made rapid progress, with the best models now rivaling human performance on the task. Here are the ExactMatch (EM) and F1 scores evaluated on the test set of v1.1.

Rank	Model	EM	F1
	Human Performance <i>Stanford University</i> (Rajpurkar et al. '16)	82.304	91.221
1	BERT (ensemble) <i>Google AI Language</i> https://arxiv.org/abs/1810.04805	87.433	93.160
2	BERT (single model) <i>Google AI Language</i> https://arxiv.org/abs/1810.04805	85.083	91.835
2	nlNet (ensemble) <i>Microsoft Research Asia</i>	85.356	91.202
2	nlNet (ensemble) <i>Microsoft Research Asia</i>	85.954	91.677
3	QANet (ensemble) <i>Google Brain & CMU</i>	84.454	90.490
4	r-net (ensemble) <i>Microsoft Research Asia</i>	84.003	90.147

SQuAD: The Stanford Question Answering Dataset

Social and Media Impact



The New York Times

News you can act on.
Reporting you can use every day.

Subscribe for
\$2 a week.

121,486 views | Dec 29, 2014, 11:37am

Tech 2015: Deep Learning And Machine Intelligence Will Eat The World



Anthony Wing Kosner Contributor

Quantum of Content and innovations in user experience

THE WALL STREET JOURNAL.

Home World U.S. Politics Economy Business Tech Markets Opinion Life & Arts Real Estate WSJ. Magazine

nature > review articles > article



nature

International journal of science

Review Article | Published: 27 May 2015

Deep learning

Yann LeCun , Yoshua Bengio & Geoffrey Hinton

Nature 521, 436–444 (28 May 2015) | Download Citation

Subscribe

BBC

Sign in

News

Sport

Weather

Shop

Reel

Travel

More

NEWS

Home | Video | World | US & Canada | UK | Business | Tech | Science | Stories | Enter

TECH

Deepfake Videos Are Getting Real and That's a Problem

Moving Upstream explores the dark side of sophisticated video fakery

By Hilke Schellmann

Oct. 15, 2018 5:29 a.m. ET

92 COMMENTS

Seeing isn't believing anymore. Deep-learning computer applications can now generate fake video and audio recordings that look strikingly real.

In a recent video published by researchers to show how the

Ads by Google

Report this ad

Why this ad?

Save On
PRINTING
Supplies!



LD COMPATIBLE T6M14AN
/ 902XL HIGH YIELD BLACK
INK FOR HP

\$16.99

Con
Si
or

SH

Science & Environment

Chasing quakes with machine learning

By Mary Halton
Science reporter, BBC News

THE SHIFT

An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy.

"I won, and I didn't break any rules," the artwork's creator says.



HA

I have a bag of apples, there are 5 apples in it, if I take 2 out of it, how many apples are left in the bag?



QA

If you have a bag with 5 apples in it and you take 2 out of it, then there will be 3 apples left in the bag. This is because $5 - 2 = 3$.



HA

If I take out 2 more apples, how many apples are still in the bag now?



QA

If you have a bag with 5 apples in it and you take 2 apples out, then there will be 3 apples left in the bag. If you take out 2 more apples, then there will be 1 apple left in the bag. This is because $3 - 2 = 1$.



In total, if you take out 4 apples from a bag with 5 apples in it, then there will be 1 apple left in the bag. This is because $5 - 4 = 1$.

HA

What if I put in 7 more?



QA

If you have a bag with 5 apples in it and you take out 4 apples, then there will be 1 apple left in the bag. If you then put in 7 more apples, then there will be 8 apples in the bag in total. This is because $1 + 7 = 8$.



HA

Write codes to get day diff from 2022-01-01 to today in JavaScript



QA

To get the number of days between a date and the current date in JavaScript, you can use the `Date` object and the `getTime()` and `getTimezoneOffset()` methods.



Here is an example of how to use the `Date` object and the `getTime()` and `getTimezoneOffset()` methods to get the number of days between a date and the current date in JavaScript:

Copy code

```
// Create Date objects for the start date and the current date
const startDate = new Date(2022, 0, 1); // January 1, 2022
const endDate = new Date();

// Get the number of milliseconds in each Date object
const startTime = startDate.getTime();
const endTime = endDate.getTime();

// Get the time zone offset in minutes for each Date object
const startOffset = startDate.getTimezoneOffset();
const endOffset = endDate.getTimezoneOffset();

// Calculate the difference in milliseconds between the two Date objects,
// taking into account the time zone offset
const diff = endTime - startTime + (endOffset - startOffset) * 60 * 1000;

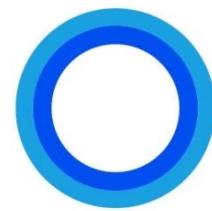
// Calculate the number of days between the two dates
const dayDiff = Math.floor(diff / (1000 * 60 * 60 * 24));

// Print the number of days between
console.log(dayDiff); // Output: (Try again) days from 2022-01-01 to today)
```

Wide Deployment in Industry



Hi, how can I help?



Hi. I'm Cortana.



Microsoft Is Looking To Add ChatGPT To Office 365...And Other Small Business Tech News This Week

Gene Marks Contributor 

I write about public policy and tech that impacts my small business.

Follow



Jan 15, 2023, 07:00am EST

Hugging Face reaches \$2 billion valuation to build the GitHub of machine learning

Romain Dillet @romaindillet / 10:28 AM EDT • May 9, 2022

 Comment

Robotics & AI

Stability AI, the startup behind Stable Diffusion, raises \$101M

Kyle Wiggers @kyle_l_wiggers / 1:01 PM EDT • October 17, 2022

 Comment

Enterprise

DeepL, the AI-based language translator, raises over \$100M at a \$1B+ valuation

Ingrid Lunden @ingridlunden / 8:20 AM EST • January 11, 2023

 Comment

Unprecedented Popularity

NeurIPS | 2018

Thirty-second Conference on Neural
Information Processing Systems



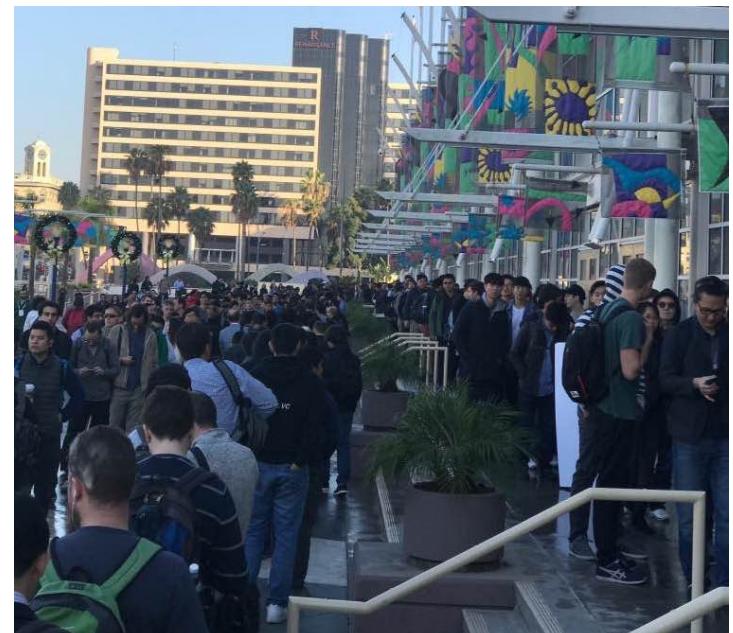
NIPS @NipsConference · 4m

#NIPS2018 The main conference sold
out in 11 minutes 38 seconds

3

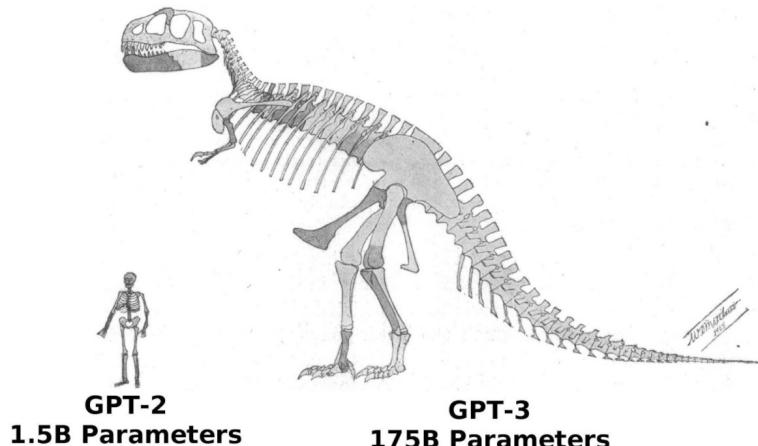
21

25



Running DNNs are expensive

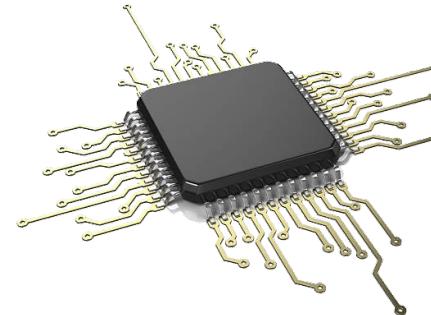
Storage intensive



175B Parameters

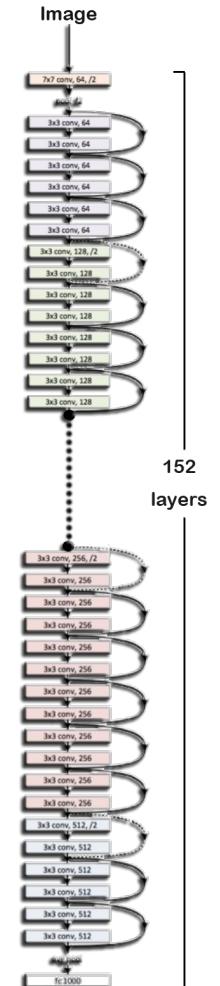
GPT-3 from OpenAI

- ## • Computation intensive



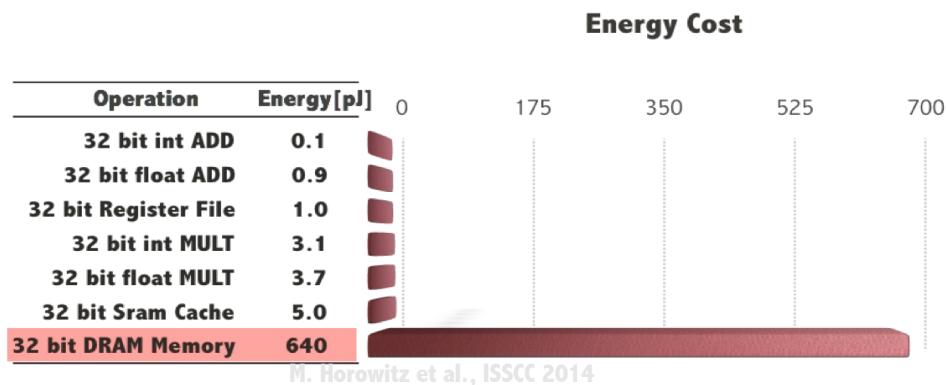
11.3G MAC

Resnet-152 from Microsoft

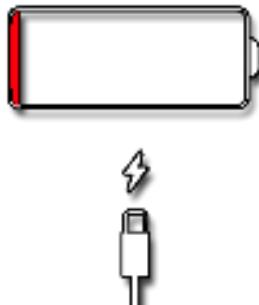


1 MAC=1 Multiplication + 1 Addition

Challenge on Computation



Mobile
Devices



Datacenter



**1920 CPUs
280 GPUs**

\$3,000/game for ⚡

Training a single AI model can emit as much carbon as five cars in their lifetimes

Deep learning has a terrible carbon footprint.

by **Karen Hao**

Energy and Policy Considerations for Deep Learning in NLP

Emma Strubell Ananya Ganesh Andrew McCallum

College of Information and Computer Sciences
University of Massachusetts Amherst

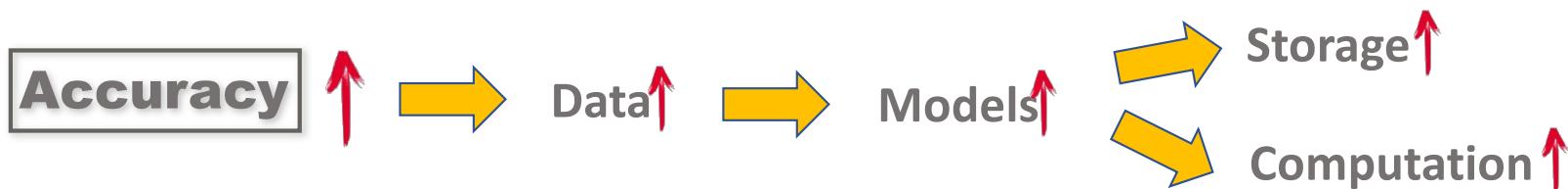
{strubell, aganesh, mccallum}@cs.umass.edu

Consumption	CO ₂ e (lbs)
Air travel, 1 passenger, NY↔SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000

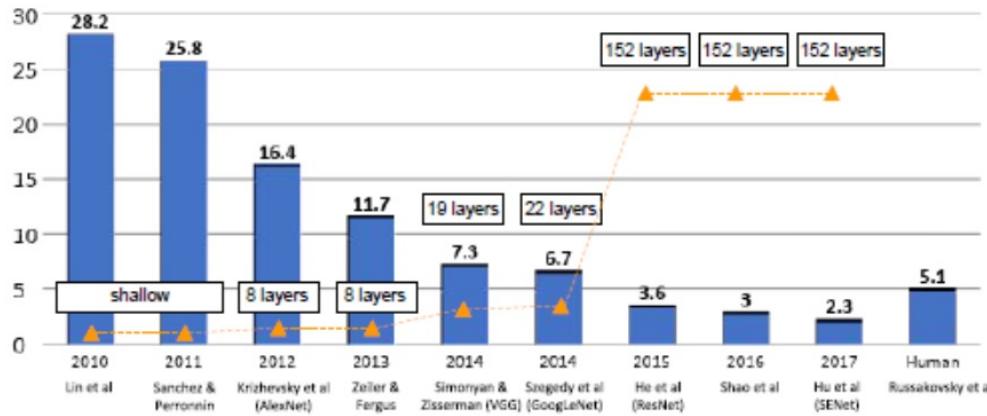
Training one model (GPU)	
NLP pipeline (parsing, SRL)	39
w/ tuning & experimentation	78,468
Transformer (big)	192
w/ neural architecture search	626,155

Table 1: Estimated CO₂ emissions from training common NLP models, compared to familiar consumption.¹

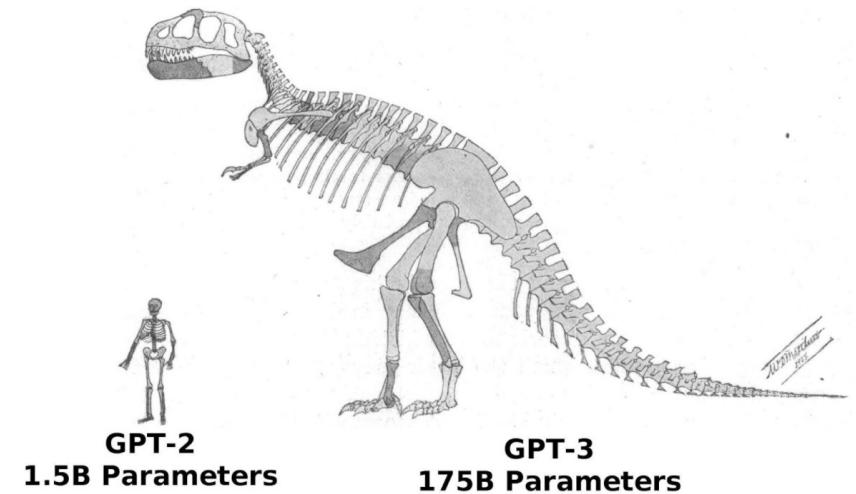
DNNs are Continuing to Scale



ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners



K. He, ICML Tutorial 2016



SWITCH TRANSFORMERS: SCALING TO TRILLION PARAMETER MODELS WITH SIMPLE AND EFFICIENT SPARSITY

William Fedus*

Google Brain

liamfedus@google.com

Barret Zoph*

Google Brain

barrettzoph@google.com

Noam Shazeer

Google Brain

noam@google.com

ABSTRACT

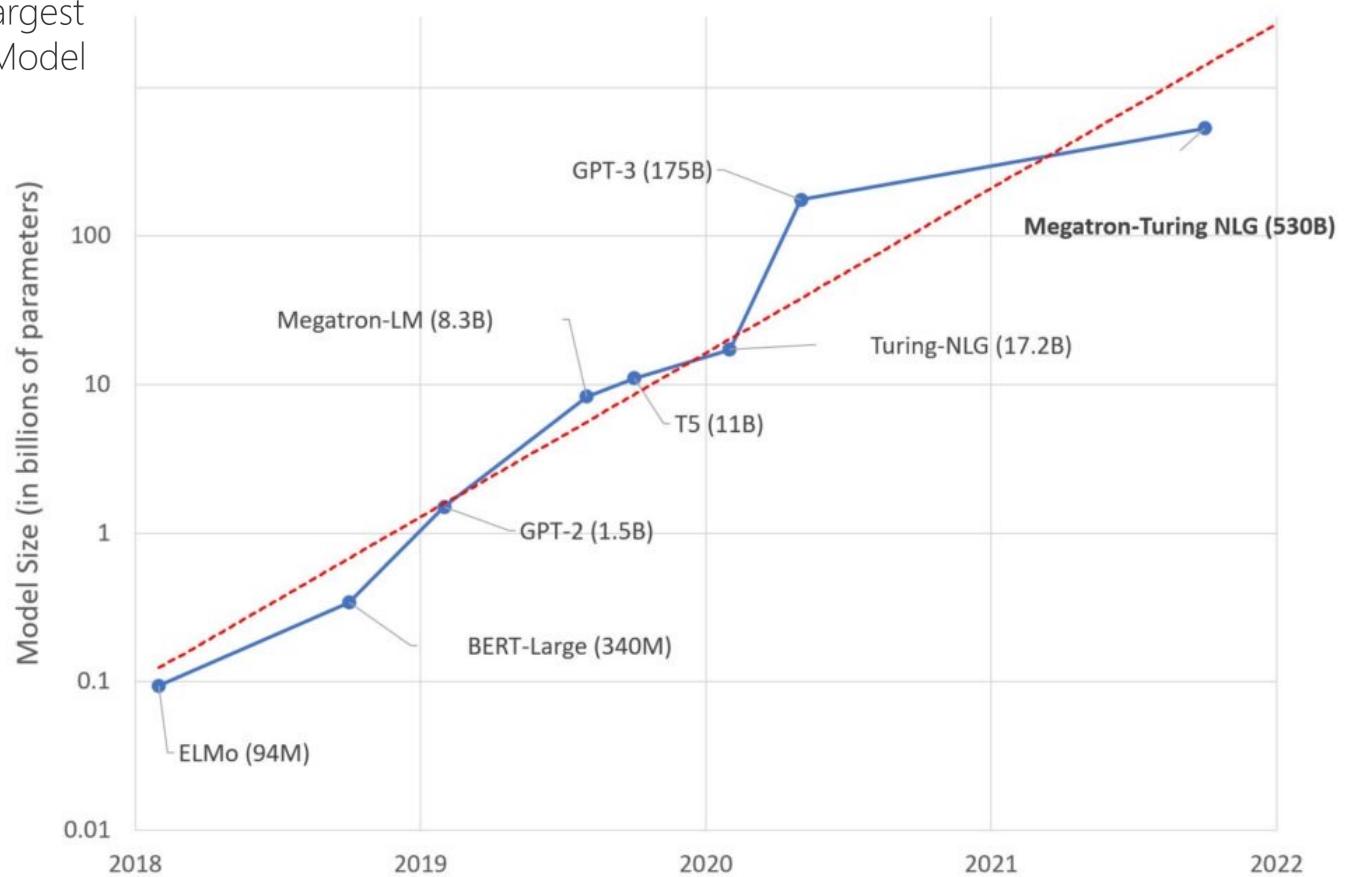
In deep learning, models typically reuse the same parameters for all inputs. Mixture of Experts (MoE) models defy this and instead select *different* parameters for each incoming example. The result is a sparsely-activated model – with an outrageous number of parameters – but a constant computational cost. However, despite several notable successes of MoE, widespread adoption has been hindered by complexity, communication costs, and training instability. We address these with the Switch Transformer. We simplify the MoE routing algorithm and design intuitive improved models with reduced communication and computational costs. Our proposed training techniques mitigate the instabilities, and we show large sparse models may be trained, for the first time, with lower precision (bf16) formats. We design models based off T5-Base and T5-Large (Raffel et al., 2019) to obtain up to 7x increases in pre-training speed with the same computational resources. These improvements extend into multilingual settings where we measure gains over the T5 Base variant across all 101 languages. Finally, we demonstrate

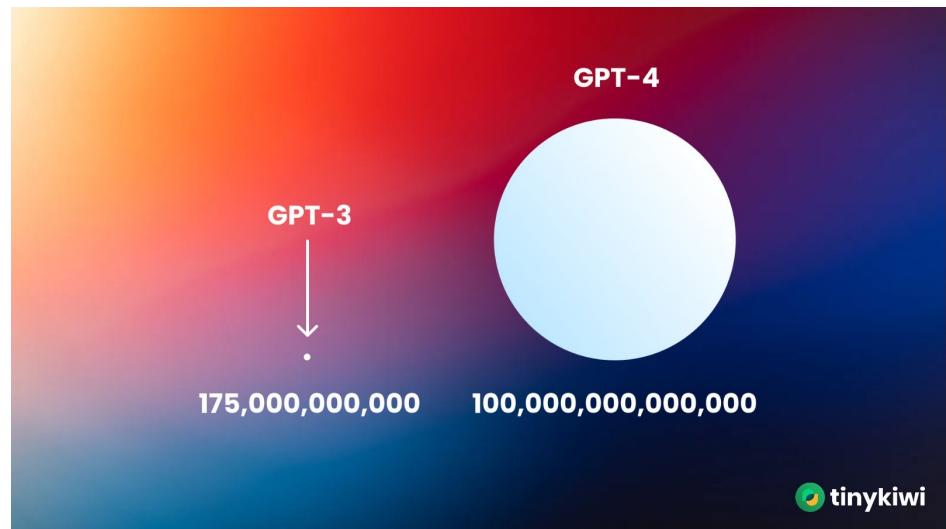
[Return to Blog Home](#)

Microsoft Research Blog

Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, the World's Largest and Most Powerful Generative Language Model

Published October 11, 2021





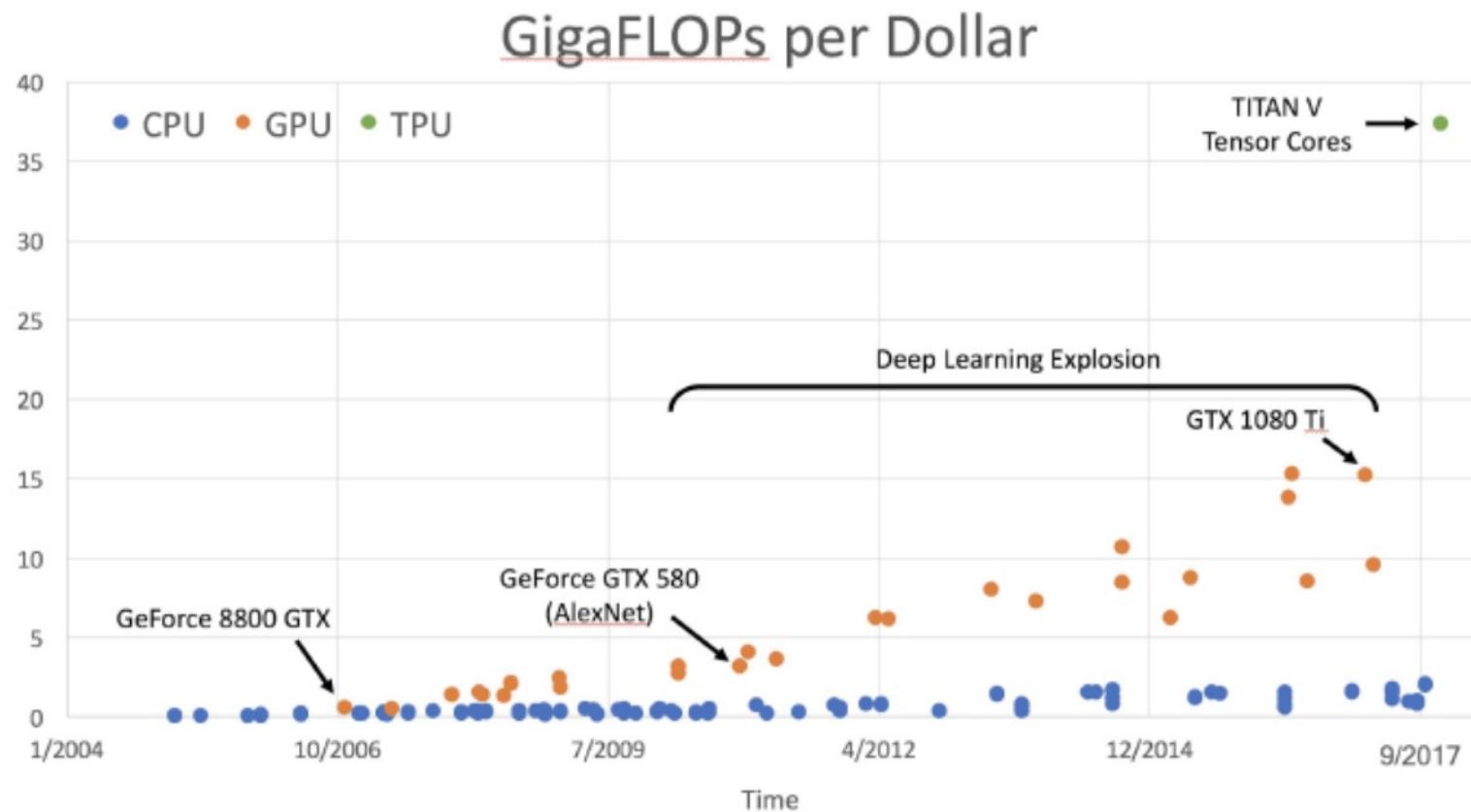
- **GPT-3:** 175 billion parameters
 - Cost (2020): \$4.6 million
- **GPT-4 (Human Brain):** 100 trillion parameters
 - Cost (2020): \$2.6 billion
 - Cost (2024): \$325 million
 - Cost (2028): \$40 million
 - Cost (2032): \$5 million

Path to Achieve Energy-efficient Deep Learning

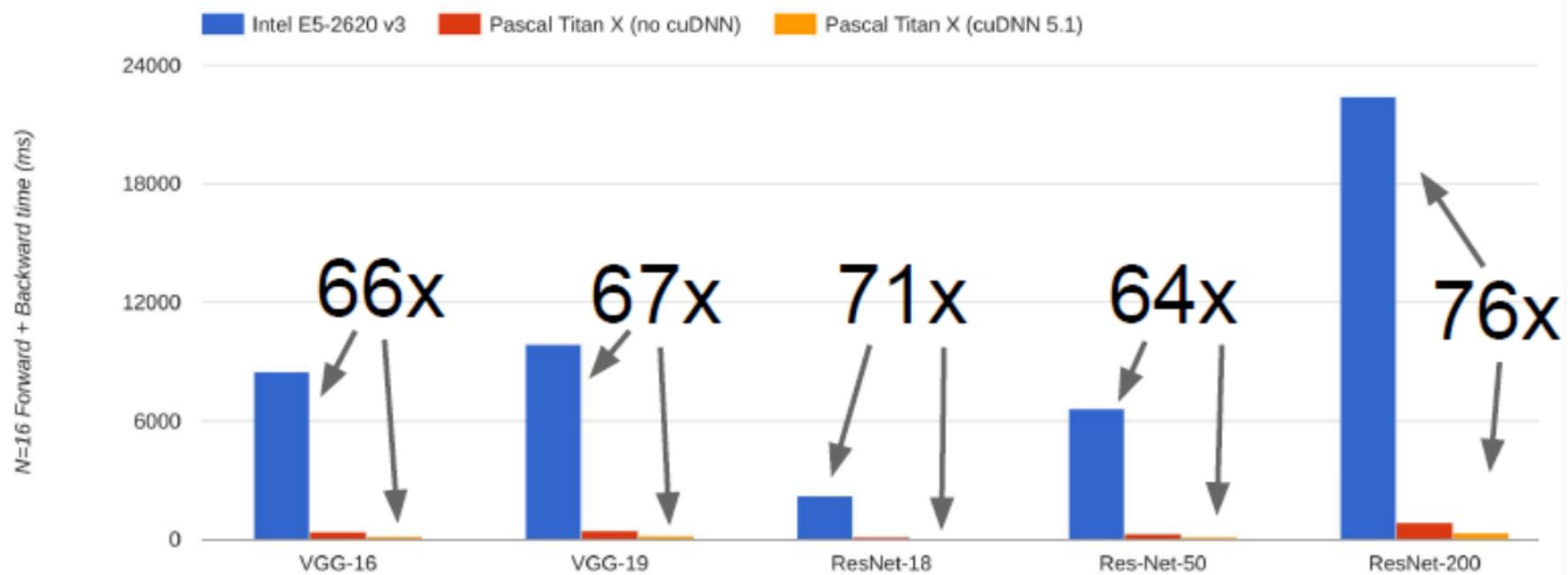
SW/HW co-design co-optimization

- Design specialized hardware architecture
 - to optimize memory access
 - to optimize computing dataflow
- Design efficient low-cost train/test algorithms
 - to compress sizes of DNNs
 - to accelerate computation of DNNs

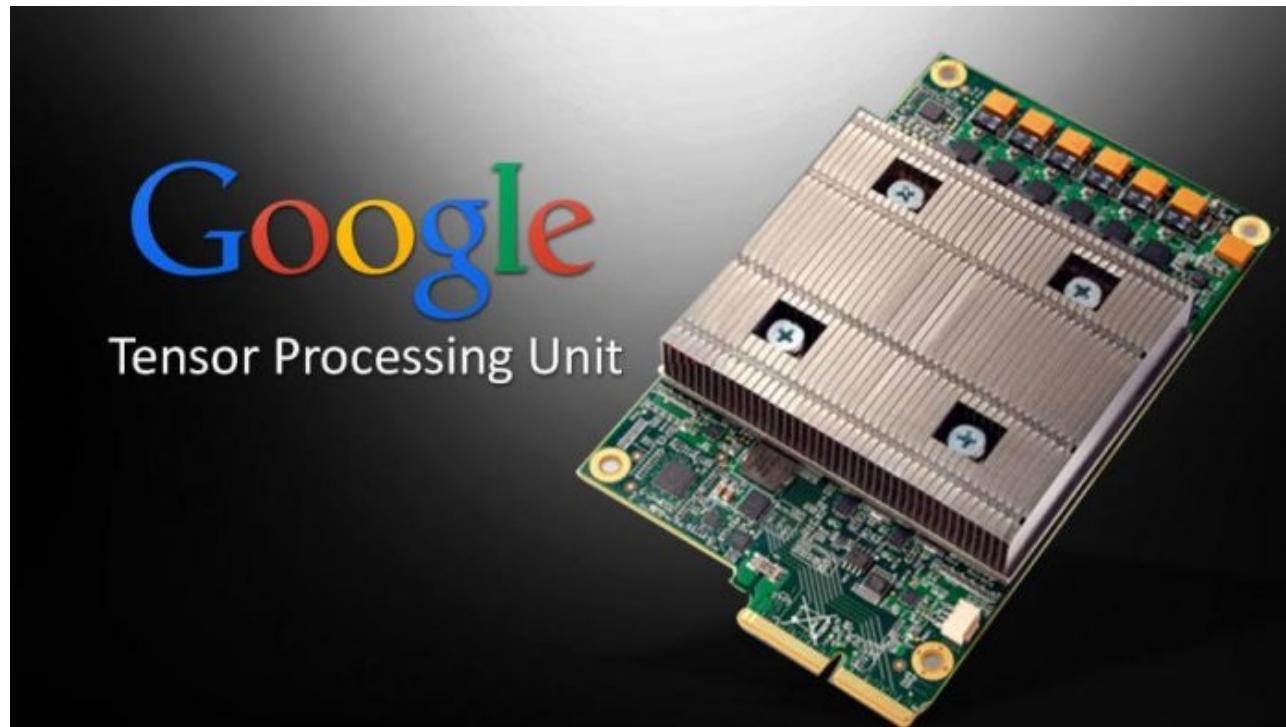
One Path: Innovation on Hardware



GPU vs CPU



Design Energy-efficient DNN is Emerging



Another Path: Innovation on Algorithm

SW/HW co-design co-optimization

- Design specialized hardware architecture
 - to optimize memory access
 - to optimize computing dataflow
- Design efficient low-cost train/test algorithms
 - to compress sizes of DNNs
 - to accelerate computation of DNNs

Covered by
This Course

[Applications](#)[Products](#)[Developers](#)[Support](#)[About](#) [\(0\)](#)

Xilinx - Adaptable. Intelligent. > News > Press > 2018 Press Releases > Xilinx Announces the Acquisition of DeePhi Tech

Xilinx Announces the Acquisition of DeePhi Tech

Deal to Accelerate Data Center and Intelligent Edge Applications

Jul 17, 2018

BEIJING and SAN JOSE, Calif., July 17, 2018 – [Xilinx, Inc.](#) (NASDAQ: XLNX) the leader in adaptive and intelligent computing, announced today that it has acquired [DeePhi Technology Co., Ltd](#) (DeePhi Tech), a Beijing-based privately held start-up with industry-leading capabilities in machine learning, specializing in deep compression, pruning, and system-level optimization for neural networks.

DeePhi Tech has been developing its machine learning solutions on Xilinx technology platforms, and the two companies have worked closely together since DeePhi Tech's inception in 2016. DeePhi Tech's neural network pruning technology has been optimized to run on Xilinx FPGAs, enabling breakthrough performance with best-in-class energy efficiency. Xilinx has been a major investor in DeePhi Tech, alongside other prominent international investors, since 2017.

"We are excited to continue our strong partnership with Xilinx and work even more closely to deliver leading machine learning solutions to our customers in China and around the world," said Song Yao, CEO of DeePhi Tech.

"Xilinx is accompanying DeePhi Tech along its journey to explore the potential of machine learning and is supporting our innovation as one of our early investors. We look forward to continuing our joint efforts with Xilinx to bring our solutions to the next level in performance," said Yi Shan, CTO of DeePhi Tech.

"We are thrilled to welcome DeePhi Tech to the Xilinx family and look forward to further building our leading engineering capabilities and enabling the adaptable and intelligent world," said Salil Raje, executive vice president of Software and IP Products Group at Xilinx. "Talent and innovation are core to realizing our vision. Xilinx will

Importance in Academia

On Device Intelligence

Search this site

ICML 2016 Workshop on On-Device Intelligence

Accepted Papers

Invited Speakers

Schedule

Organizers

Important Dates

Call for Contributions

ICML 2016 Workshop on On-Device Intelligence

NIPS 2018 CDNNRIA Workshop

Compact Deep Neural Network Representation with Industrial Applications

 Montreal, Canada  December 3-8, 2018  <https://nips.cc/Conferences/2018/Schedule?showEvent=109...>

This workshop aims to bring together researchers, educators, practitioners who are interested in techniques as well as applications of making compact and efficient neural network representations.

One main theme of the workshop discussion is to build up consensus in this rapidly developed field, and in particular, to establish close connection between researchers in machine learning community and engineers in industry. We believe the workshop is beneficial to both academic researchers as well as industrial practitioners.

tinyML Foundation

Enabling ultra-low Power Machine Learning at the Edge

[Asia 2021](#)[Summit 2022](#)[Research Symposium 2022](#)[All Events](#)

All Events



Research Symposium 2022

Mar 28, 2022

[More About](#)

Summit 2022

Mar 28, 2022 - Mar 30, 2022

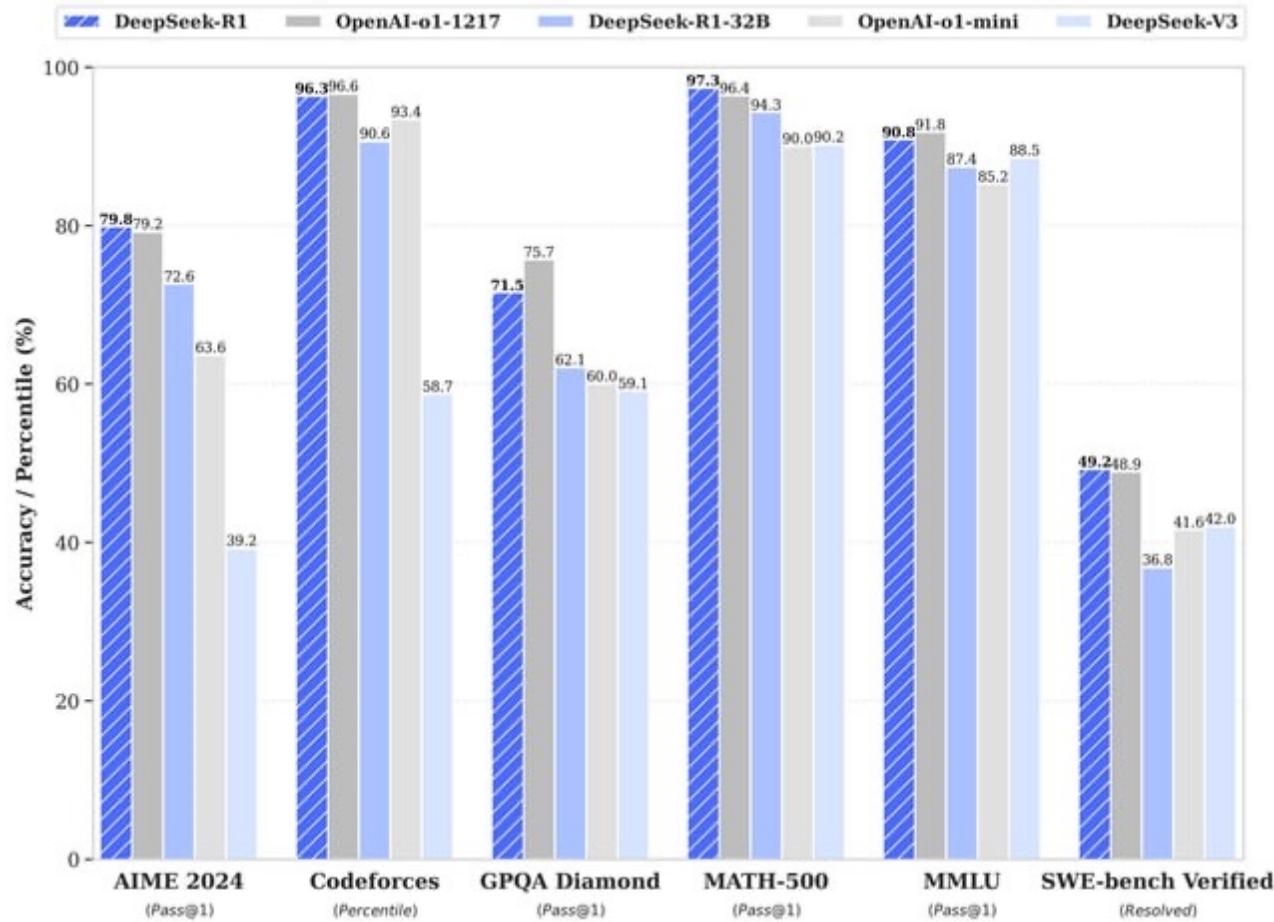
[More About](#)

tinyML Talks: Enabling Low Power
ML at the Edge-tinyML Shenzhen
Kick-off Networking Meetup

Jan 21, 2022

[More About](#)

DeepSeek-R1 uses pure reinforcement learning to match OpenAI o1 – at 95% less cost

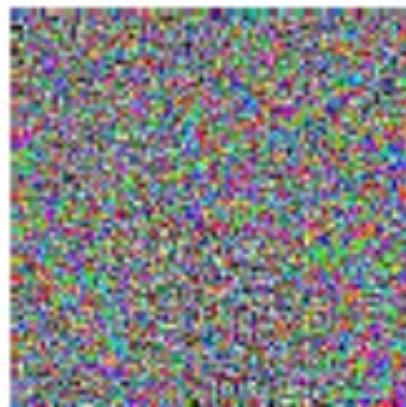


Security of Deep Learning



x
“panda”
57.7% confidence

$+ .007 \times$



$\text{sign}(\nabla_x J(\theta, x, y))$
“nematode”
8.2% confidence

$=$



$x +$
 $\epsilon \text{sign}(\nabla_x J(\theta, x, y))$
“gibbon”
99.3 % confidence

“EXPLAINING AND HARNESSING ADVERSARIAL EXAMPLES,” I. Goodfellow etc.

Three Small Stickers in Intersection Can Cause Tesla Autopilot to Swerve Into Wrong Lane

Security researchers from Tencent have demonstrated a way to use physical attacks to spoof Tesla's autopilot

By Evan Ackerman



Fig 29. Left picture shows we add some noise on the left lane line in digital level, and right picture shows the result of APE's lane recognition function. (We redact top left of our image for privacy reasons, but it won't affect the final result.)



All images: Tencent

Course Information

Topics Covered in This Course

- Machine learning basics
- DNN basics, construction, training, application
- Algorithm design for DNN
- Vulnerability and Robustness of DNN

Expected Learning Outcomes

- Knowledge of math foundation of DNN
- Coding experience for DNN training/evaluation
- Knowledge of compact and fast DNN design
- Knowledge for attack & defending DNN models

Related Courses

- 01:198:444 Topics in Computer Science (Deep Learning)
- 16:198:536 Machine Learning
- 16:198:534 Computer Vision
- 16:198:520 Introduction to Artificial Intelligence
- 16:198:535 Pattern Recognition: Theory & Applications
- 16:198:543 Massive Data Storage, Retrieval and Deep Learning
- 16:332:561 Machine Vision
- 16:332:579 Machine Learning for IoT
- 16:332:445 Machine Learning for Engineers
- 16:332:512 Nonlinear Adaptive Control and Learning
- 16:332:525 Optimum Signal Processing and ML
- 16:332:539 High-dimensional models in DSP and ML
- 16:332:549 Detection and Estimation Theory: Inference and ML

Related Conferences

CSRankings: Computer Science Rankings

CSRankings is a metrics-based ranking of top computer science institutions around the world. Click on a triangle (▶) to expand areas or institutions. Click on a name to go to a faculty member's home page. Click on a chart icon (the after a name or institution) to see the distribution of their publication areas as a bar chart. Click on a Google Scholar icon (✉) to see publications, and click on the DBLP logo (✉) to go to a DBLP entry. Applying to grad school? Read this first. Do you find CSrankings useful? Sponsor CSrankings on GitHub.

Rank institutions in by publications from to

All Areas [\[off | on\]](#)

AI [\[off | on\]](#)

▼ Artificial intelligence

ACM SIGAI, AAAI

AAAI

IJCAI

▼ Computer vision

CVF

CVPR

ECCV

ICCV

▼ Machine learning

ACM SIGKDD, ICLR, IMLS,
NEURIPS/NIPS

ICLR

ICML

NeurIPS/NIPS

KDD

► Natural language processing

- 1 ▶ Carnegie Mellon University
- 2 ▶ Univ. of Illinois at Urbana-Champaign
- 3 ▶ Georgia Institute of Technology
- 3 ▶ Massachusetts Institute of Technology
- 5 ▶ Univ. of California - San Diego
- 5 ▶ University of Michigan
- 5 ▶ University of Washington
- 8 ▶ Cornell University
- 8 ▶ Stanford University
- 10 ▶ Purdue University
- 10 ▶ Univ. of California - Berkeley
- 10 ▶ Univ. of California - Los Angeles
- 10 ▶ University of Maryland - College Park
- 14 ▶ Columbia University
- 14 ▶ New York University
- 14 ▶ Northeastern University
- 14 ▶ Rutgers University

<http://csrankings.org>

Office Hours

- Instructor
 - Time: Tuesday. 4:00pm-6:00pm, CORE 719
(or appointment via email)

Teaching Assistant

- Zhihao Tan
- Office hour for TA: Wed. 1:00-3:00pm, EE 216

Communication

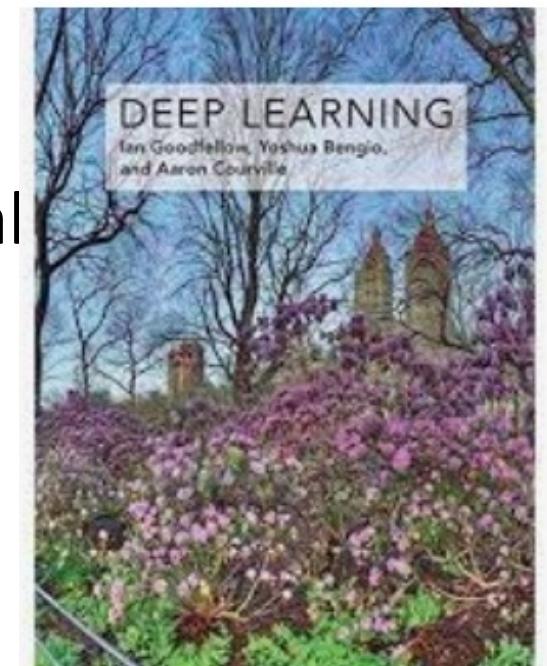
A lot of information will be posted on the Canvas

- Syllabus
- Lecture slides
- Homework & solutions
- Reading materials: papers, tutorials
- Any update & notification

Check your Canvas inbox frequently

Course Materials

- No textbook required (DL develops fast)
- Reference book
 - “Deep Learning” by Ivan Goodfellow et.al
- Online tutorials links (very useful)
- Research papers
- Online courses (see acknowledgement)



This Course is Very “Diverse”

- Course content spans software and hardware
- Background of class vary very much
- Students have different computing resource (CPU/GPU)

Diverse Background of Class

- 40+ undergraduate students
- 60+ graduate students
- Home department/program: EE, CE, CS, ME

Grading

- Homework (50%)
- Midterm/Quizzes (20%)
- Final Project (30%)

Homework

- A typical homework may contain handwritten and coding
 - Submit online via Canvas
- No late homework (unless for special circumstances)
- 10 % penalty per calendar day for late homework
- No late homework once solutions are posted on Canvas

Quiz/Midterm

- We will have either (TBD)
 - One midterm (in person)
 - Two quizzes (online)
- No makeup midterm/quizzes for any reason

Final Programming Project

- Two types of project: Type-A and Type-B
- Type-A project
 - task to use small-size neural network on small dataset
 - CPU is enough for Type-A project
- Type-B project
 - task to use large-size neural network on large dataset
 - GPU is preferred for Type-B project
- You can choose both Type-A and Type-B projects
 - Only one type of project (higher score) is needed for grading

Type-A Project

- The task is fixed and given by instructor
- Some candidate tasks can be chosen
- A project report is required at the end of semester
- No team for Type-A project
- Recommended for students who have little prior experience in DNNs or only CPU resource
- Details will be posted on Canvas

Potential Type-A Project Tasks

- Some potential Type-A project tasks
 - manually write a small CNN on a small dataset for image classification for target accuracy
 - write a small RNN/LSTM on a small dataset for speech recognition or language translation for target accuracy
- Here “Manually” means no library module, no auto gradient

Type-B Project

- The task is flexible
 - Some candidate tasks will be given
 - Students can also propose your own projects
- A project report is required at the end of semester
- Team (2 students/per team) allowed
- Recommended for students who have some prior experience in DNNs and GPU resource

Propose Your Own Type-B Projects

- Good opportunity to explore some advanced topics
 - Replicate experiment of a published paper
 - Apply DNNs for a domain problem (CV, NLP etc.)
 - A new idea to improve performance of DNN
- A project proposal is needed for approval
 - Background, motivation, plan, target...
 - Complexity should be higher than Type-A project

Potential Type-B Project Tasks

- Some potential Type-B project tasks
 - Applying large DNNs in classical tasks in CV, NLP etc.
 - Applying large DNNs in new domains in EECS and more
 - Replicate a DNN attack/defense approach
 - Replicating a DNN pruning method on large network
 - ...
- You can propose your own idea

Programming Languages

- Python is mandatory
 - No C/C++, Matlab, R, or Java any more...
- Recommend Python
- Links to Python online tutorials will be posted in Canvas

Why Python

- Dominated in machine learning community
- Very easy to use, well-suited for beginners
- The two most popular deep learning frameworks are based on python

Zoo of Deep Learning Frameworks

Caffe



TensorFlow

theano

mxnet

The torch logo features a teal molecular structure icon followed by the word "torch".

PYTORCH

The PaddlePaddle logo includes a stylized paddle icon followed by the text "PaddlePaddle".



Deep Learning Frameworks

- Deep learning frameworks can be viewed as “library” specific for neural networks.
- If not forbidden, using frameworks can save you a lot of time in programming
- Pytorch is recommended in this course
- A short tutorial of installing Pytorch are posted in Canvas
- You are free to use your preferred framework
- TAs’ support on Pytorch, Tensorflow

Why Pytorch

- Very easy for beginners to use
- Very popular in both academia and industry

Criteria	Winner	Runner up
Community and Support	Tensorflow	Pytorch
Ease of Use	Pytorch	Tensorflow
Academia(Prototyping)	Pytorch	Torch/Caffe
Industry	Tensorflow	Caffe2
Embedded Computer vision	Caffe	Tensorflow

<https://cv-tricks.com/>



Andrej Karpathy

@karpathy

Following

Matlab is so 2012. Caffe is so 2013. Theano is so 2014. Torch is so 2015. TensorFlow is so 2016. :D

RETWEETS
218

LIKES
590



12:08 PM - 8 Feb 2017

45

218

590



Andrej Karpathy

@karpathy

Following

I've been using PyTorch a few months now and I've never felt better. I have more energy. My skin is clearer. My eye sight has improved.

1:56 AM - 26 May 2017

384 Retweets 1,519 Likes



33



384



1.5K



Posted by u/Wild_Quiet8627 7 months ago 



[D] Google quietly moving its products from Tensorflow to JAX

Discussion

<https://www.businessinsider.com/facebook-pytorch-beat-google-tensorflow-jax-meta-ai-2022-6>

With companies and researchers leaving Tensorflow and going to PyTorch, Google seems to be interested in moving its products to JAX, addressing some pain points from Tensorflow like the complexity of API, and complexity to train in custom chips like TPU. The article says that JAX still has long way to go since it lacks proper optimization to GPUs and CPUs when compared to TPUs.

 125 Comments

 Share

 Save

 Hide

 Report

94% Upvoted

Some Suggestion to Learn Deep Learning and This Course

- Understand the underlying math
 - Math is always important
- Spend time on programming and model tuning
 - Deep learning is famous for its “black box” property
- Search, Read, and Try
 - Github, Arxiv, Stack Overflow, Reddit, Hugging Face...
- Practice, Practice, and Practice

Acknowledgement

Many materials of the slides of this course are adopted and re-produced from several deep learning courses and tutorials.

- Prof. Fei-fei Li, Stanford, CS231n: Convolutional Neural Networks for Visual Recognition (online available)
- Prof. Andrew Ng, Stanford, CS230: Deep learning (online available)
- Prof. Yanzhi Wang, Northeastern, EECE7390: Advance in deep learning
- Prof. Jianting Zhang, CUNY, CSc G0815 High-Performance Machine Learning: Systems and Applications
- Prof. Vivienne Sze, MIT, “Tutorial on Hardware Architectures for Deep Neural Networks”
- Pytorch official tutorial <https://pytorch.org/tutorials/>