Introduction

Deep Learning

What is deep learning?

- Deep learning is a branch of machine learning involving algorithms that have many nonlinear processing stages.
- In most cases, deep learning refers to training neural networks that have many layers.
- A network with no hidden layer can only create linear functions or linear decision boundaries.
- A network with only one hidden layer (a shallow network) can create any practical function or decision boundary.
- This is why researchers before 2005 did not focus deeper networks, although training algorithms already existed.
- Once deep networks became popular, the numbers of layers increased rapidly (hundreds of layers).
- Now research is focusing on more sophisticated and efficient architectures

What spurred deep learning - New Initialization methods

- Many deep learning concepts were known in the 1990s.
- Deep networks were tested in the 1990s, but shallow networks tended to perform just as well.
- Deep network training was difficult because of two main problems:
  - Vanishing gradient
  - Exploding gradient
- In 2006 three different research groups (Hinton, Bengio, LeCun) published methods that could be used to initialize deep networks so that training became practical.
- It was later found that these initialization techniques were not required, but they initiated a new look at deep networks.

What spurred deep learning - GPUs

- In 2006, Nvidia introduced the cuda language for multi-purpose programming of GPUs.
- Several groups then started experimenting with GPUs for neural network training.
- In 2009 a group led by Andrew Ng used GPUs to train a deep belief network and achieved 75x speedup over CPU implementations.
- In 2010 a group led by Jurgen Schmidhuber trained a deep multilayer network on GPUs and achieved the highest accuracy on the benchmark MNIST data set of handwritten digits.
- In 2012 a group led by Geoffrey Hinton won the Imagenet competion by a large margin with a convolution network trained on GPUs.
- The networks that won these competitions were multilayer or convolution networks that were introduced in the 1980s.

What spurred deep learning - large data sets

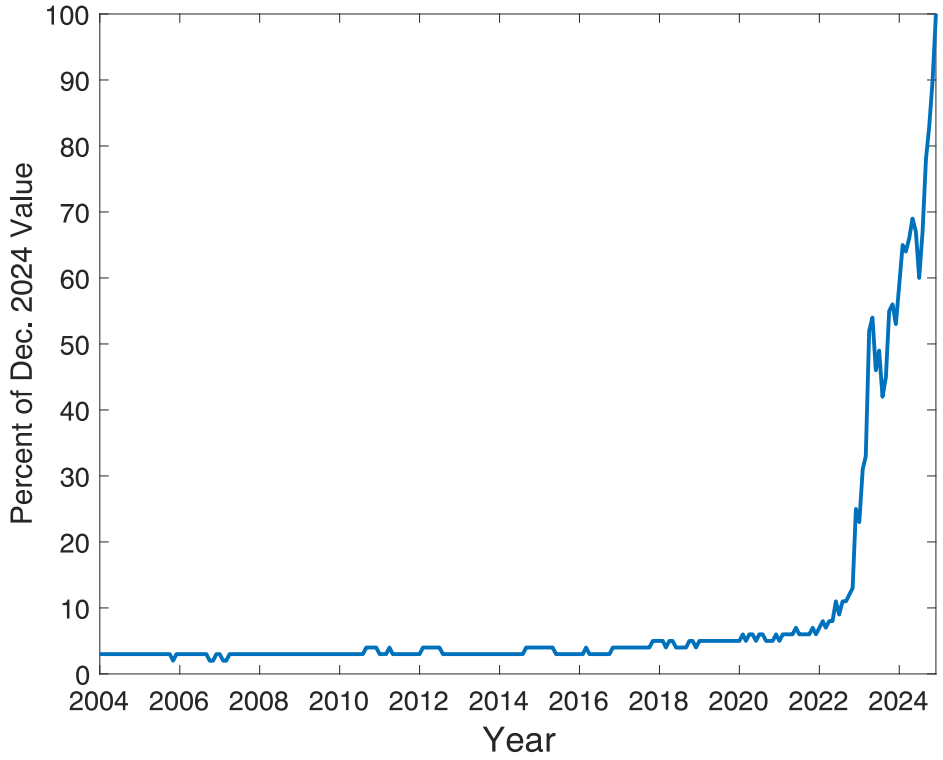
- Practical deep networks can have hundreds of millions of weights (the largest can have many billions).
- It requires large data sets to train these networks.
- During the 2010s and beyond many large data sets began to be collected.
- ImageNet is one of the most popular image databases with more than 14 million hand-annotated images.
- CIFAR-10 & CIFAR-100 are labeled subsets of the 80 million tiny images dataset.
- The MS COCO (Microsoft Common Objects in Context) dataset has 328K images.
- IMDB-Wiki contains 523,051 images in total, with 460,723 face images from 20,284 celebrities from IMDb and 62,328 from Wikipedia.
- ArXiv: a dataset of all the papers on ArXiv as fulltext, with over 270 GB of data.

What spurred deep learning - open source frameworks

- The earliest deep network development required cuda coding.
- After 2006, deep learning frameworks were developed, with higher level languages interfacing to cuda libraries.
- In 2007 the University of Montreal introduced Theano, with a Python interface.
- The Torch framework came in 2002 (École Polytechnique Fédérale de Lausanne), using the Lua language. By 2007, Torch7 included cuda libraries. It was converted to Python by Meta – PyTorch (2017).
- Caffe was developed at the University of California, Berkely in 2013. It had Matlab and Python interfaces, but was mainly accessed by JSON files.
- TensorFlow has a Python interface and was developed by the Google Brain group starting in 2011. It was released as TensorFlow open source in 2015.

What spurred deep learning - more developers

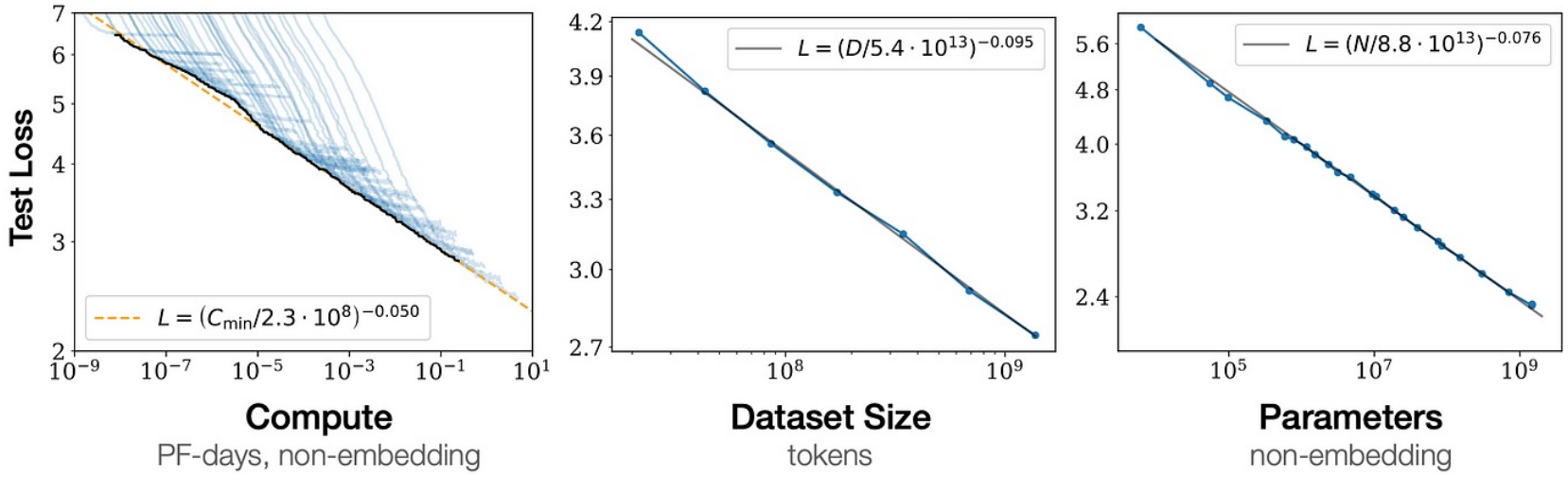
- A final factor that has accelerated deep learning is the sheer number of researchers in the field.
- When ChatGPT was launched in November of 2022, it received 152 million visitors in its first month.



Google searches for AI from 2004 through 2024

The Scaling Law

- A key idea driving deep learning is the scaling law.
- As computing power, data set size and model size increase, the performance will increase also.
- The corollary is that if these do not continue to increase, progress may become more difficult.



From [Kaplan et al., 2020])

Applications

- There have been many successful applications of neural networks since the late 1980s.
- Progress on applications continued throughout the 1990s and 2000s.
- With deep learning, performance for the most complex problems has improved tremendously.
- Especially in computer vision and natural language processing, deep networks are exponentially better.
- The Nobel prizes in both Chemistry and Physics in 2024 were awarded to deep learning researchers.

AlphaFold

- AlphaFold is a deep learning system developed by DeepMind.
- Predicts the three-dimensional structures of proteins based on amino acid sequences.
- One of the most significant advancements in the field of computational biology.
- As accurate as X-ray crystallography and cryo-electron microscopy.
- Predicted structures for almost all proteins in the human proteome.
- Won the Nobel prize in Chemistry.

Large Language Models (LLMs)

- LLMs are deep neural networks that are trained to predict the next word in a sequence of text.
- Current LLMs use the transformer architecture, invented in 2017.
- LLMs are used in all chatbots, like ChatGPT and Claude.
- Revolutionized the way that humans interact with computers.
- Have passed the Uniform Bar Exam and the U.S. Medical Licensing Examination.
- Can write computer code from a natural language description of the desired operation.

GANs and Diffusion for Image Generation

- Generative Adversarial Networks (GANs) were first introduced in 2014.
- Diffusion models were introduced in 2015.
- Together, they enabled neural network generation of artificial images (deep fakes).
- Photo-realistic images can be created from a simple text description.
- Short films have been generated using this technology.
- Actors and screenwriters have gone on strike to prevent the use of this technology.

Speech Generation and Recognition

- 2016 – WaveNet (DeepMind) is first neural network to generate realistic human-like speech as raw audio.
- 2023 – VALL-E 2 (Microsoft) can generate high-quality personalized speech using only a three-second recording.
- 2016 – Deep Speech (Mozilla/Baidu) was an end-to-end deep learning model that could recognize either English or Mandarin Chinese.
- 2020 – Wav2Vec 2.0 (Facebook) broke the record for accuracy on a key speech recognition benchmark with 100x less labeled data
- 2023 – Whisper (OpenAI) works on languages it has never seen; can convert spoken words from one language into text in another.