

Deep Learning Crash Course



www.deeplearningcrashcourse.org

Hui Xue

Fall 2022



National Heart, Lung,
and Blood Institute

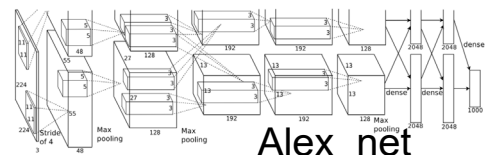
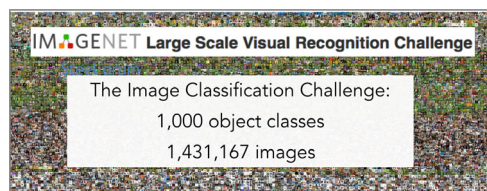
What we want to achieve

- Introduce the basics of deep learning
- Present in-depth how DL model works
- Provide practices to build your own model
- Grow interest and improve community awareness



After this course and assignments, start to apply DL to your research ...

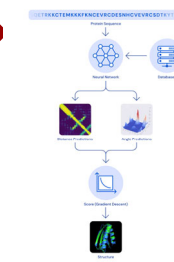
<https://arxiv.org/abs/1112.6209>



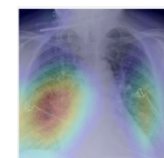
<https://github.com/cs231n/cs231n.github.io>
<https://dl.acm.org/doi/10.1145/3065386>



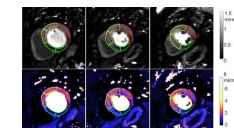
<https://www.youtube.com/watch?v=tIThdr3O5Qo>



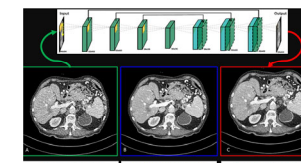
<https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery>



<https://pubs.rsna.org/doi/10.1148/rvaj.2021190228>



<https://pubs.rsna.org/doi/10.1148/rvaj.2020200009>



<https://www.sciencedirect.com/science/article/pii/S1120179720302866?via%3Dihub>



<https://medium.com/vsinghbisen/how-ai-based-drone-works-artificial-intelligence-drone-use-cases-7f3d44b8abe3>

NHLBI Deep Learning Crash course aims to provide individuals with the ability to apply these tools to their own research.

What is Deep Learning?

Deep Learning is a computational tool that trains neural networks to detect and infer information on an input signal. With this, the system is able to learn from the data and make predictions that are often better than the actual conclusion of the input. In the application of this research, a Deep Learning can also be seen as an application of machine learning. Machine learning is a branch of computer science that focuses on the development of algorithms that can learn from data and make predictions or decisions based on the results of the learning and execution of some specific tasks.

What are the Big Issues with Deep Learning Technology?

In their article and video lectures, Google and IBM/DJAI researchers discuss the challenges and issues of deep learning. Tasks in the context of various models such as gradient descent and reinforcement learning. They also address concerns about the lack of interpretability of deep learning models and the need for attention on the problem of output recognition and explanation, as well as the need for more research on the

Text Generation API

The text generation API is backed by a large scale unsupervised language model that generates paragraphs of text. This transformer-based language model, based on the GPT-2 model by OpenAI, intakes a sentence or partial sentence and predicts subsequent

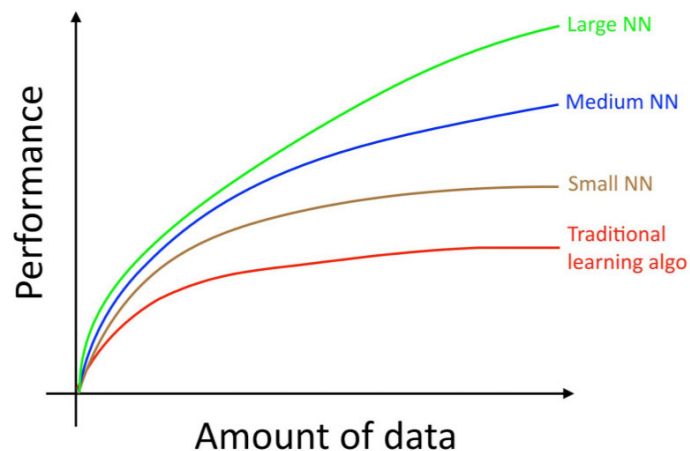
<https://deepai.org/machine-learning-model/text-generator>

NIH National Heart, Lung,
and Blood Institute

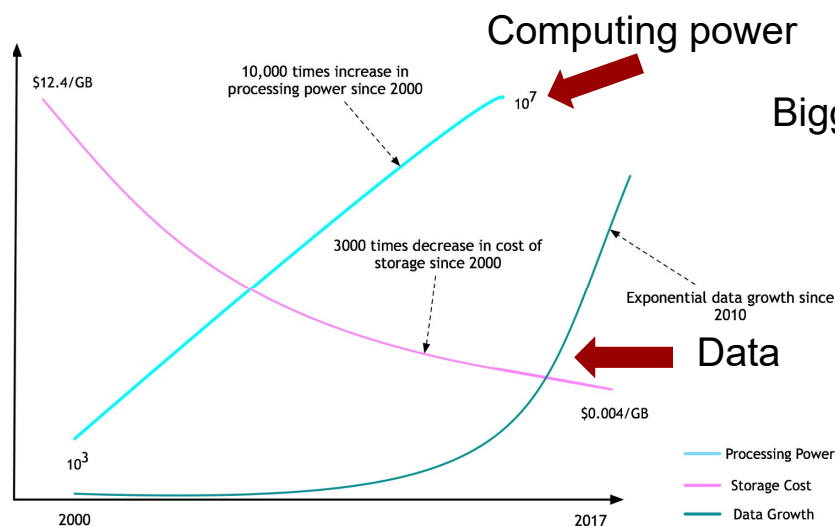
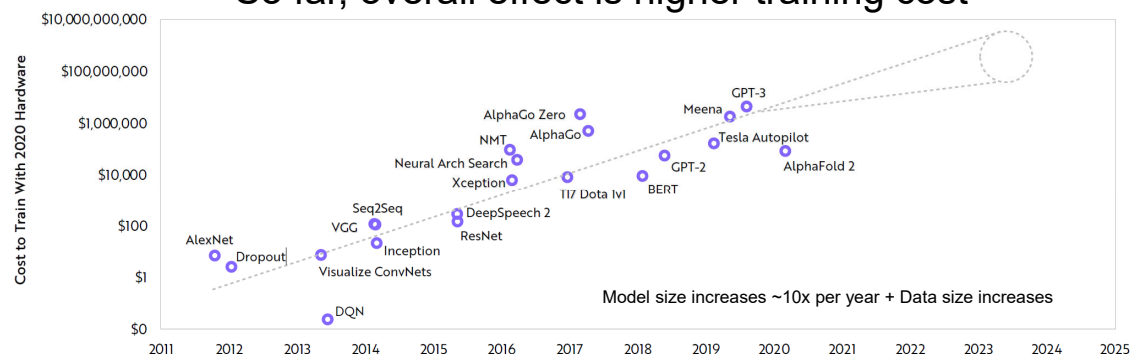
3

A combination of Data, Computing, Models

So far, overall effect is higher training cost

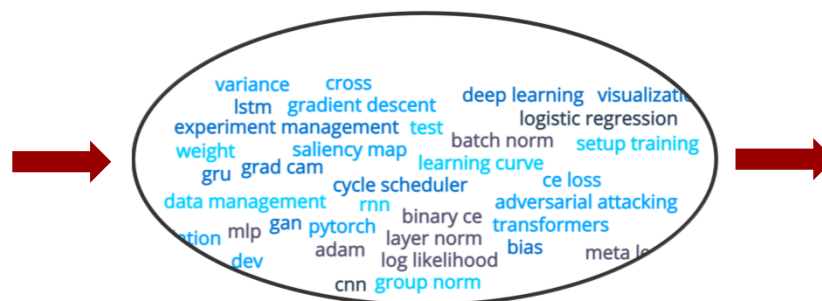
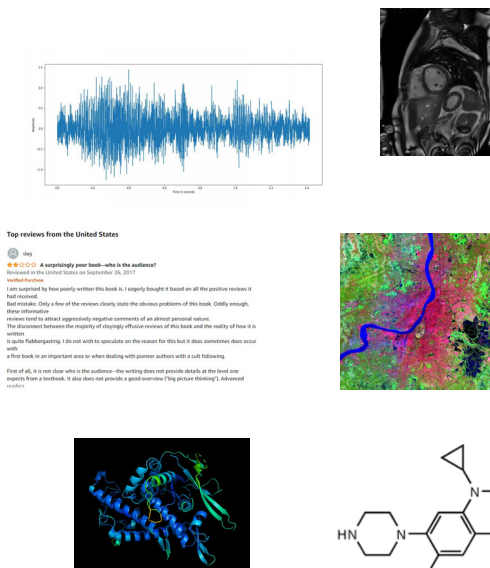


Why deep model succeeds?



Why learn DL: same technology, widely applicable

Multi-modality data

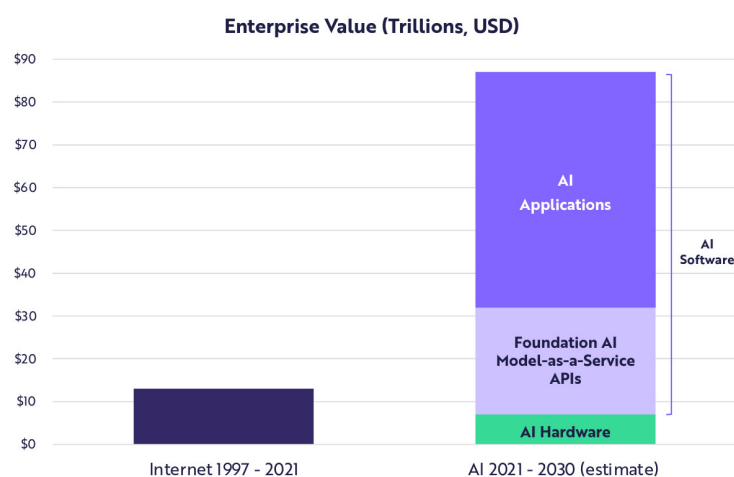


We will learn what is in the
deep learning toolbox

- Automation
- AI assistant
- Prediction
- High duration system with 0% down time
- Super-human performance in some applications
- ... still rapidly evolving

<https://www.nature.com/articles/d41586-020-03348-4>
<https://towardsdatascience.com/review-deep-learning-in-drug-discovery-f4c89e3321e1>

Benefits forecast



Research forecasts big contribution from AI tech to the global economy.

-- Big Idea 2022, <https://ark-invest.com/big-ideas-2022/>

AI ADOPTION by INDUSTRY and FUNCTION, 2021

Source: McKinsey & Company, 2021 | Chart: 2022 AI Index Report

| | Human Resources | Manufacturing | Marketing and Sales | Product and/or Service Development | Risk | Service Operations | Strategy and Corporate Finance | Supply-chain Management |
|--|-----------------|---------------|---------------------|------------------------------------|------|--------------------|--------------------------------|-------------------------|
| All Industries | 9% | 12% | 20% | 23% | 13% | 25% | 9% | 13% |
| Automotive and Assembly | 11% | 26% | 20% | 15% | 4% | 18% | 6% | 17% |
| Business, Legal, and Professional Services | 14% | 8% | 28% | 15% | 13% | 26% | 8% | 13% |
| Consumer Goods/Retail | 2% | 18% | 22% | 17% | 1% | 15% | 4% | 18% |
| Financial Services | 10% | 4% | 24% | 20% | 32% | 40% | 13% | 8% |
| Healthcare Systems/Pharma and Medical Products | 9% | 11% | 14% | 29% | 13% | 17% | 12% | 9% |
| High Tech/Telecom | 12% | 11% | 28% | 45% | 16% | 34% | 10% | 16% |

% of Respondents (Function)

Figure 4.3.2

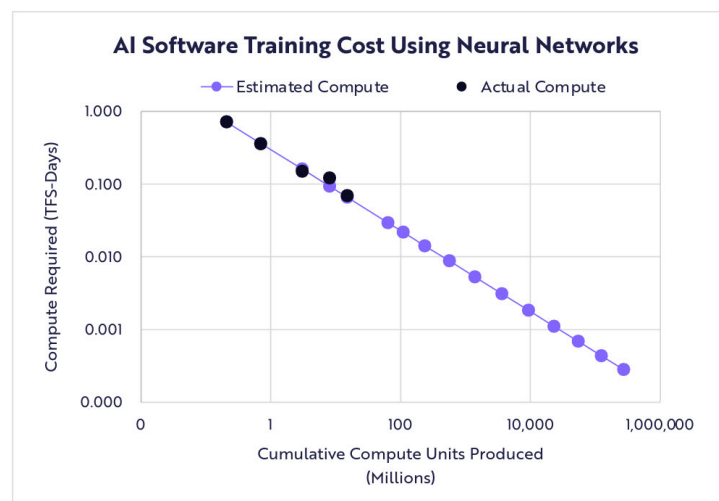
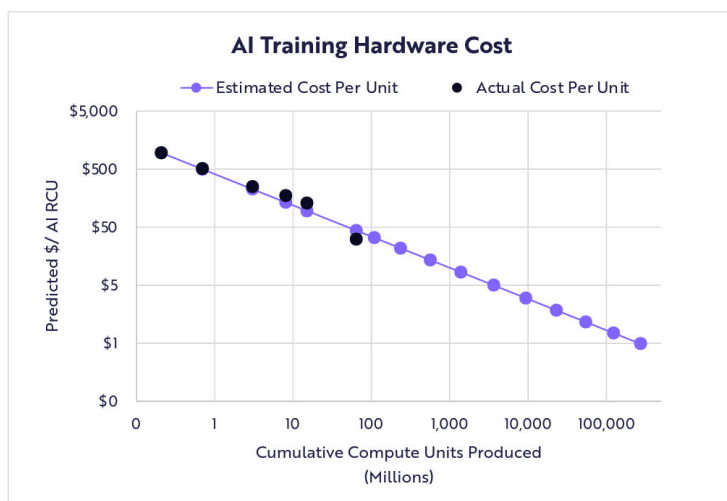
https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf

- Since 2018, the cost to train an image classification system has decreased by 63.6%, while training times have improved by 94.4%.
- Adoption of AI technique increases from ~16% to ~25%, since 2020
- 1,843 participants, survey

Cost decreases

Wrights law: Theodore Paul Wright, also known as T. P. Wright, While studying airplane manufacturing, Wright determined that for **every doubling of airplane production** the labor requirement was reduced by 10-15%.

The cost of each unit produced decreases as a function of the cumulative number of units produced.



Combined hardware and software progress could drive down AI training costs by 60% at an annual rate by 2030 -> more applications and more development

For 2022 Fall offering

- Course logistics

Information and Course Resources

- Website – www.deeplearningcrashcourse.org
- Assignment – 3 in total
- Office hours – Friday tutorial, demo, Q&A, discussion
- Course instructors:



Hui Xue (hui.xue@nih.gov)

More information

- For this offering – www.deeplearningcrashcourse.org/nhlbi2022
 - Detailed introduction for every lecture
 - Reading list
- Information for Setup

https://deeplearningcrashcourse.org/setup_nhlbi/

We will use NHLBI windows computer for Part 1. Everyone should be able to do it.

https://deeplearningcrashcourse.org/setup_nhlbi_linux/

There are also linux servers prepared to use. **It will be still available to you after the course.**

- Require to know some python programming
 - Will demo some basics and how to debug the code

Two parts

- Deep Learning Fundamentals consists of 7 lectures and one session

This fall!

Neural network, training, optimization, backprop, CNN and variants, applications ...

- Deep learning Advances consists of 5-6 lectures

Next spring

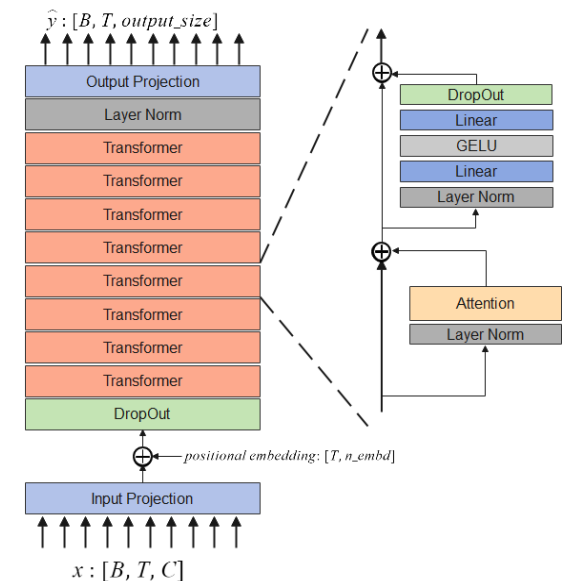
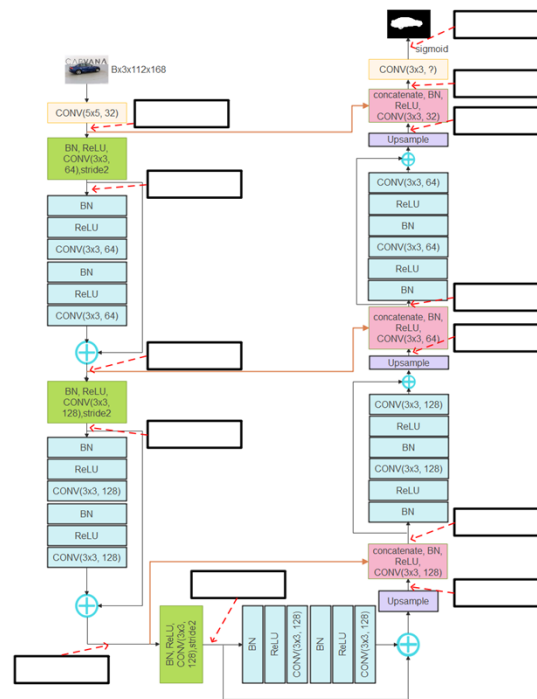
More advanced topics (transformer, GAN, visualization etc.)

Assignments

■ assignments

| | |
|-----------|---|
| A1 | Neural Network basics, Multi-layer Perceptron, Gradient descent |
| A2 | Backprop, Hyperparameter searching, Setup training, Pytorch |
| A3 | CNN, model training, Segmentation |

- Many coding problems
- Tooling for testing, experimental management, hyper-parameter searching ...
- Based on feedback, we will release solutions and go through problems in the Q&A sessions together



From you

Deep learning and neural networks is becoming an essential in the field of Big Data and Biomedicine. I want to develop on my machine learning skills to stay relevant.

I am interested in deep learning because the many possible applications to my research. I use all-atom and coarse-grained models to understand the aggregation charged lipids and the interactions of peptide-based drugs with lipoproteins. Deep learning is a possible tool for improving all-atom and coarse-grained models, characterizing current simulations, and predicting future peptide-based drugs to investigate. My goal is to directly apply what I learn in the course to my on-going research.

My postdoctoral project involves using deep learning to segment cellular organelles in platinum replica EM images of human cells. Over the last year, I have worked extensively with the NIH AIM core (Jiamin Liu) to test and implement a machine learning model to identify and segment organelles in a dense EM image. This course is directly related to my postdoctoral project and will help me gain a deep and extensive understanding of modern deep learning tools for applications to image segmentation, denoising, and analysis. I have extensive experience with coding (matlab and python) for modeling and image analysis.

I am running MD simulations that generate very large datasets and my current analysis on these datasets is limited to about three dimensions. I am interested in using deep learning to identify important dynamics in my simulations, improve the parameters for MD simulations, and to guide enhanced sampling methods.

I am intending to apply deep learning to the massive genomic and epigenomic data generated in our lab.

I am highly interested in this course. I am currently in a master program in AI, and would love to supplement my studies.

I would like to learn more about deep learning and its implementation, and apply the methods to my statistical and collaborative research.

General interest of the theory and the tools so we can support staff in DIR.

The recent advances in protein structure prediction have piqued my interest in Deep Learning. I would like to study this topic in detail to better understand how it can be applied to my work.

Currently, I am interested in learning about and using bioinformatics and data science-based techniques to investigate the structural, functional, and genetic underpinnings of hematological disorders, with a focus on platelets. I would like to learn about deep learning and neural networks since both are being applied to great success in analyzing vast amounts of image data in platelet spreading assays and other platelet-focused methods (e.g. Kempster et al. 2022: doi: 10.1038/s41598-022-08613-2)

My work towards structural determination of proteins in cryo-electron micrographs necessitates a working knowledge of state-of-art image analysis, where methods in deep learning plays a foundational role. I'm interested in learning both how to use deep learning in my research and how to evaluate novel deep learning implementations in the field.

I'd like to understand the theoretical background behind deep learning models so that I can build my own models or choose existing models effectively. The primary applications would be noise reduction and prediction of multiple protein folds from one sequence.

I am interested in deep learning in relation to the signal processing side of biological spectroscopy. Particularly for data smoothing, data compression for denoising, or automated classification methods that could reduce some of our manual processing steps.

I spent the past semester at Brown learning the basics of Numpy and Python. My research interests lay with the dynamics of protein interactions during exocytosis and using this information to generate models of membrane fusion and fission. I am a new graduate student and this course will give me the foundation to apply machine/deep learning to my thesis project which will involve a great deal of imaging and image analysis with both light and electron microscopy.

I am a fellow at NHLBI, NIH and PhD student at University of Maryland. I am working on the selectivity mechanism of Topoisomerases and I use Molecular Dynamics simulations to aid the single molecule experiments. I often get a lot of data from simulation and I try to look at the patterns and their correlation with the experiment results. And I feel that knowing deep learning technique would significantly help my quest to find patterns



National Heart, Lung,
and Blood Institute