CIE 555
Neural Networks and Deep Learning
Course Introduction

Dr. Mohamed Elshenawy
melshenawy@zewailcity.edu.eg
Zewail University of Science and Technology

1

Overview

- Course Administration
- What is an Deep learning?
- History of Deep learning
- Why Deep Learning?

2

Course Administration

- **Instructor:** Mohamed Elshenawy
- **TA:** Yara Alaa
- **Email:** melshenawy@zewailcity.edu.eg
- **Office:** Room (S-021) **-** Nano Building (Second floor)
- **Office hour:** Tuesdays from 10:30 to 12:30
- **Lectures:** Sunday 3:30 PM - 5:20 PM;   F025-NB
- **Practice Session:** Monday 12:30 AM - 03:15 PM   R001-NB

جامعة زويل للعلوم والتكنولوجيا
Zewail University of Science and Technology

3

Communications

- Course web page on google classroom
  - Slides, supplementary material, grades, etc.
  - Announcements
  - Information about your labs, assignments, etc.
- Emails and office hours.
- Self-service for final grades

جامعة زويل للعلوم والتكنولوجيا
Zewail University of Science and Technology

4

Course objectives

After completing this course, you should be able to:

- Understand the fundamentals of neural networks and how they are used to solve machine learning problems.
- Identify key deep learning models and their applications in areas such as computer vision, speech recognition, robotics and natural language processing.
- Describe common optimization strategies in training deep architectures.
- Design from scratch and train deep convolutional and recurrent neural networks models.
- Understand selected deep unsupervised learning techniques and outline some of their applications
- Discuss some of the ongoing deep learning research efforts and recognize some of the open problems

جامعة زويل للعلوم والتكنولوجيا
Zewail University of Science and Technology

5

Recommended Background

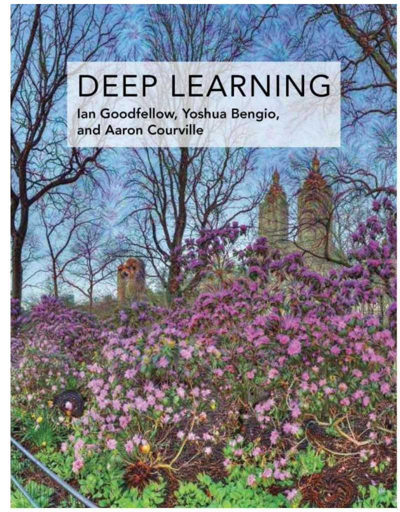
- Introduction to Machine Learning (CIE 417)
  - Common Regression and classification techniques;
  - parameter estimation techniques (e.g. least squares, Maximum likelihood);
  - optimization methods (gradient descent);
  - loss function formulation;
  - Overfitting; and
  - cross validation
- Knowledge of Python programming language
- Knowledge of linear algebra, calculus, and basic probability theory.

جامعة زويل للعلوم والتكنولوجيا
Zewail University of Science and Technology

6

Textbook

- Goodfellow, I., Bengio, Y., & Courville, A.(2016). Deep learning (Vol. 1). Cambridge: MIT press.



جامعة زويل للعلوم والتكنولوجيا
Zewail University of Science and Technology

7

Grading

- Two progress exams (20%)
- Project (25%)
- Final Exam (40%)
- Class participation and attendance (5%)
- Assignments and Exercises (10%)

Students are expected to recognize and uphold standards of intellectual and academic integrity. Plagiarism, fabrication and cheating will not be tolerated. To know about the university of Science and Technology's code of conduct, please check the university website https://www.zewailcity.edu.eg

75% attendance is required. No Makeup for the progress exams.

جامعة زويل للعلوم والتكنولوجيا
Zewail University of Science and Technology

8

Lectures and topics

- Introduction to Deep Learning.
- Deep Feedforward Networks
- Regularization Methods.
- Optimization Methods for Training Deep Models.
- Convolutional Neural Networks.
- Recurrent Neural Networks and Sequence Modeling.
- Autoencoders.
- Deep Generative Models.

Zewail University of Science and Technology

9

Neural Networks and Deep Learning: An introduction

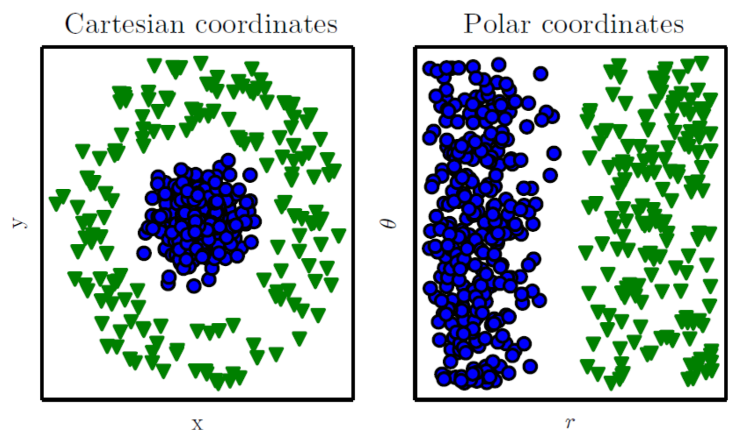
Chapter 1

Zewail University of Science and Technology

10

Representations Matter

- For many tasks, it is difficult to know what features should be extracted
- Many artificial intelligence tasks can be solved if we choose the right set of features and how they are represented



Goodfellow 2016

11

Feature Extraction

- Suppose that we would like to write a program to detect cars in photos, which features can be used?
- When designing features or algorithms for learning features, our goal is usually to separate the **factors of variation** that explain the observed data.
- Such factors are often not quantities that are directly observed.
- They can be thought of as concepts or abstractions that help us make sense of the rich variability in the data (e.g. the position of the car, its color, and the angle and brightness of the sun).
- It is usually difficult to extract such high-level, abstract features from raw data.

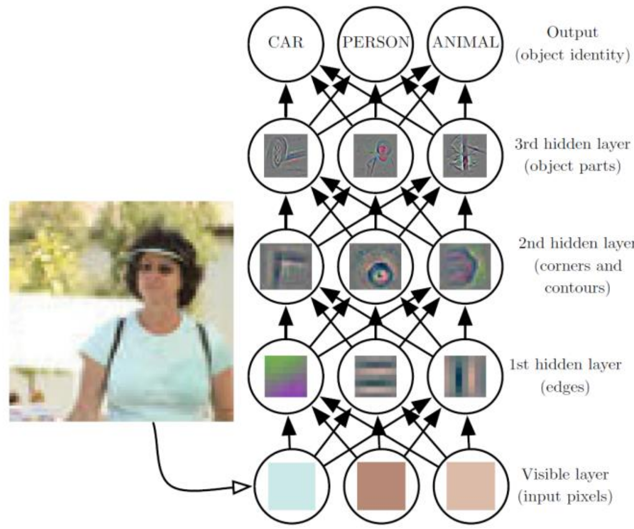
Goodfellow 2016

Zewail City
Zewail University of Science and Technology

12

Deep Learning – learning feature representations

- **How do we learn useful representations?**
- **Deep learning** solves this problem by introducing representations that are expressed in terms of other, simpler representations.
- It allows the computer to build complex concepts out of simpler concepts.
- a deep learning system can represent the concept of an image of a person by combining simpler concepts, such as corners and contours, which are in turn defined in terms of edges.



Goodfellow 2016

13

Another perspective

- Another perspective on deep learning is that depth allows the computer to learn a multi-step computer program.
- Each layer of the representation can be thought of as the state of the computer's memory after executing another set of instructions in parallel.
- Networks with greater depth can execute more instructions in sequence.
- According to this view of deep learning, not all of the information in a layer's activations necessarily encodes factors of variation that explain the input.
- The representation stores state information that helps to execute a program that can make sense of the input.
- It has nothing to do with the content of the input specifically, but it helps the model to organize its processing.

Goodfellow 2016

14

What is deep learning?

- There is no consensus about how much depth a model requires to qualify as "deep."
- Can be defined as the study of models that either involve a greater amount of composition of learned functions or learned concepts than traditional machine learning does.
- The concept can be applied in machine learning frameworks that are not necessarily neural-ly inspired.
- Neuroscience is regarded as an important source of inspiration for deep learning researchers, but it is no longer the predominant guide for the field.

Goodfellow 2016

15

Historical trends in Deep Learning

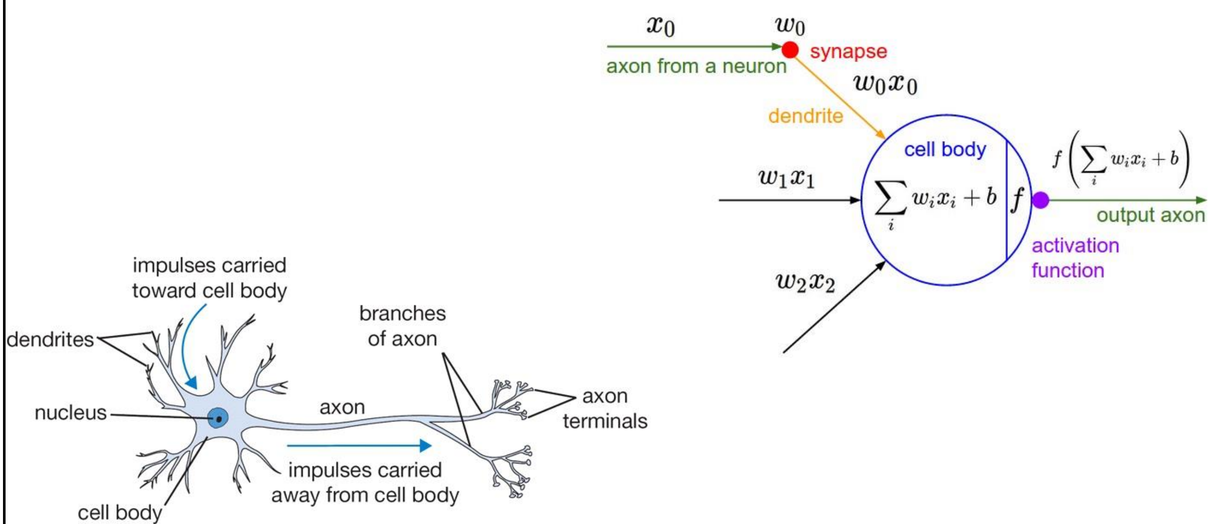
- deep learning dates back to the 1940s.
- Broadly speaking, there have been three waves of development of deep learning:
  - **cybernetics** in the 1940s–1960s
  - **connectionism** in the 1980s–1990s
  - The current resurgence under the name deep learning beginning in 2006.

- First Generation Neural Networks (McCulloch and Pitts, 1943)- proposed a highly simplified computational model of the neuron.
- This model could recognize two different categories of inputs by testing whether $f(x, w)$ is positive or negative.

Goodfellow 2016

16

Neural Network Model



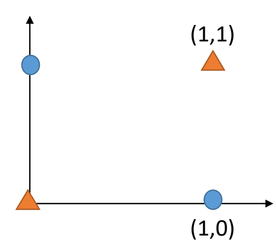
http://cs231n.github.io/neural-networks-1/

Zewail University of Science and Technology

17

Perceptron and ADALINE

- In the 1950s, the perceptron (Rosenblatt, 1958, 1962) became the first model that could learn the weights defining the categories given examples of inputs from each category. Another parameter learning technique is the **adaptive linear element** (ADALINE), which dates from about the same time.
- Models based on the $f(x,w)$ used by the perceptron and ADALINE are called **linear models**.
- Linear models have many limitations. Most famously, they cannot learn the XOR function



Goodfellow 2016

Zewail University of Science and Technology

18

The second wave of neural network research (**Connectionism**)

- The central idea in connectionism is that a large number of simple computational units can achieve intelligent behavior when networked together.
- Several key concepts arose during the connectionism movement of the 1980s remain central to today's deep learning:
  - **Distributed representation:** the idea that each input to a system should be represented by many features, and each feature should be involved in the representation of many possible inputs.
  - Using **back-propagation** to train deep neural networks.
  - Hochreiter and Schmidhuber (1997) introduced the **long short-term memory** or **LSTM** network to resolve some of these difficulties.
- Unrealistically ambitious claims and the advances in other fields of machine learning led to a decline in the popularity of neural networks that lasted until 2007.

Goodfellow 2016

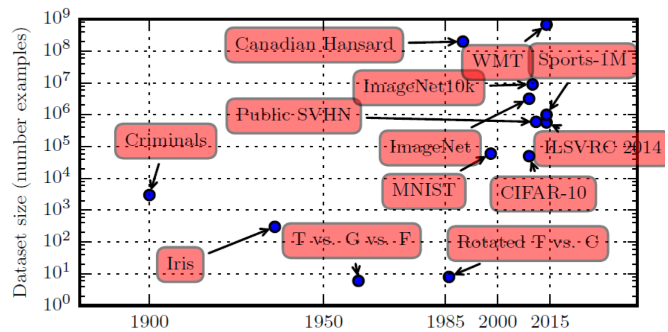
The third wave of neural networks

- began with a breakthrough in 2006. Geoffrey Hinton showed that a kind of neural network called a deep belief network could be efficiently trained using a strategy called greedy layer-wise pre-training.
- The term "deep learning" was used to emphasize that researchers were able to train deeper neural networks, and to focus their attention on the theoretical importance of depth.
- deep neural networks outperformed competing AI systems based on other machine learning technologies.

Goodfellow 2016

Why deep learning algorithms?

- **Increasing Dataset Sizes:** today we can provide DL algorithms with the resources they need to succeed (driven by the increasing digitization of society)
- As of 2016, a rough rule of thumb is that a supervised deep learning algorithm will generally achieve acceptable performance with around 5,000 labeled examples per category,
- and will match or exceed human performance when trained with a dataset containing at least 10 million labeled examples.



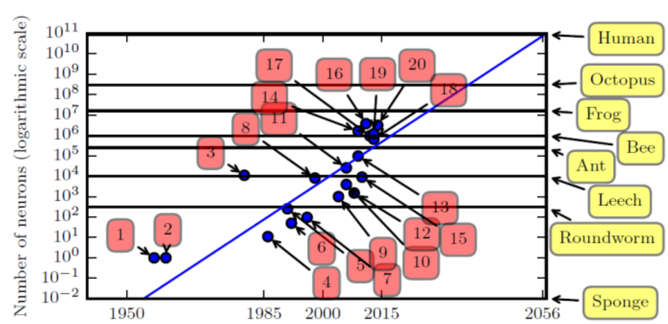
Goodfellow 2016

Zewait University of Science and Technology

21

Why deep learning algorithms?

- **Increasing Model Sizes:** we have the computational resources to run much larger models today. (driven by faster computers with larger memory and by the availability of larger datasets.)
- Unless new technologies allow faster scaling, artificial neural networks will not have the same number of neurons as the human brain until at least the 2050s.



Goodfellow 2016

Zewait University of Science and Technology

22

Why deep learning algorithms?

- **Increasing Accuracy, Complexity and Real-World Impact**
- The largest contest in object recognition is the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) held each year.
- A dramatic moment in the meteoric rise of deep learning came when a convolutional network won this challenge for the first time and by a wide margin, bringing down the state-of-the-art top-5 error rate from 26.1% to 15.3% (Krizhevsky, Sutskever, and Hinton, University of Toronto,2012)
- Since then, these competitions are consistently won by deep convolutional nets, and as of this writing, advances in deep learning have brought the latest top-5 error rate in this contest down to 3.6%

Goodfellow 2016

جامعة زويل للعلوم والتكنولوجيا
Zewail University of Science and Technology

23

ImageNet (2012)

- The goal of this competition is to estimate the content of photographs for the purpose of retrieval and automatic annotation using a subset of the large hand-labeled ImageNet dataset (10,000,000 labeled images depicting 10,000+ object categories) as training.
- The validation and test data for this competition will consist of 150,000 photographs, collected from flickr and other search engines, hand labeled with the presence or absence of 1000 object categories.

IMAGENET

http://image-net.org/challenges/LSVRC/2012/index

جامعة زويل للعلوم والتكنولوجيا
Zewail University of Science and Technology

24

ImageNet top-five classification error (%) evolution

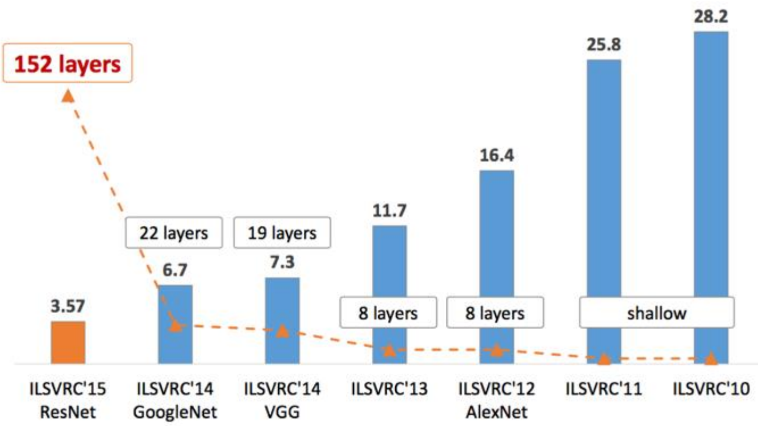


Image is taken from the Kaiming He presentation, *Deep Residual Learning for Image Recognition*

جامعة زويل للعلوم والتكنولوجيا Zewail University of Science and Technology

25

Machine Learning -Review

- "A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P* , if its performance at tasks in *T* , as measured by *P* , improves with experience *E*." Mitchell (1997)
- **Tasks:** described in terms of how the machine learning system should process an **example** (a collection of features that have been quantitatively measured from some object or event).
- Example tasks: classification, regression, transcription, machine translation, anomaly detection, etc.
- **Performance:** specific to the *task T* (e.g. accuracy, error rate)
- **The Experience, *E*:** Most machine learning algorithms simply experience a dataset.

Goodfellow, Bengio, Courville 2016

جامعة زويل للعلوم والتكنولوجيا Zewail University of Science and Technology

26

Next Lecture

- Capacity, Overfitting and Underfitting
- Cross-Validation
- Estimators, Bias and Variance
- Maximum likelihood estimation
- Maximum A Posteriori (MAP) Estimation

Zewail University of Science and Technology