Introduction to Deep Learning

Jinchao Xu

PennState

2019.7.8

Course info

Introduction to Deep Learning

Math 497, Summer 2019

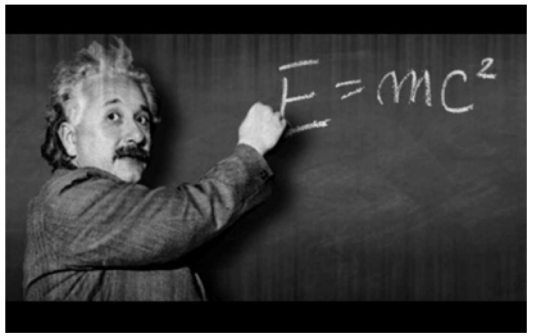
Time: 8-9:30am and 2-4pm
Every weekday

Classroom: Room 408, Third Teaching Building
(三教408)

Instructor: Jinchao Xu
Office: 1271s
Email: xu@math.psu.edu
Office hours: 10-11:30am, Tuesday and Thursday

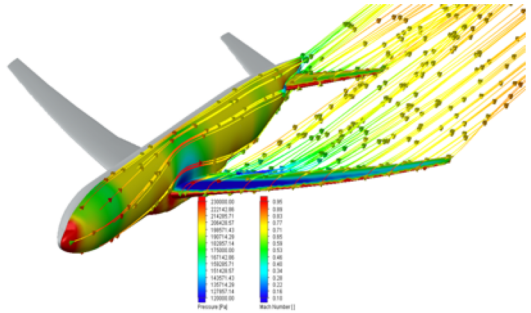
Four major methods of scientific research

Experimental science



Theoretical science

Computational science



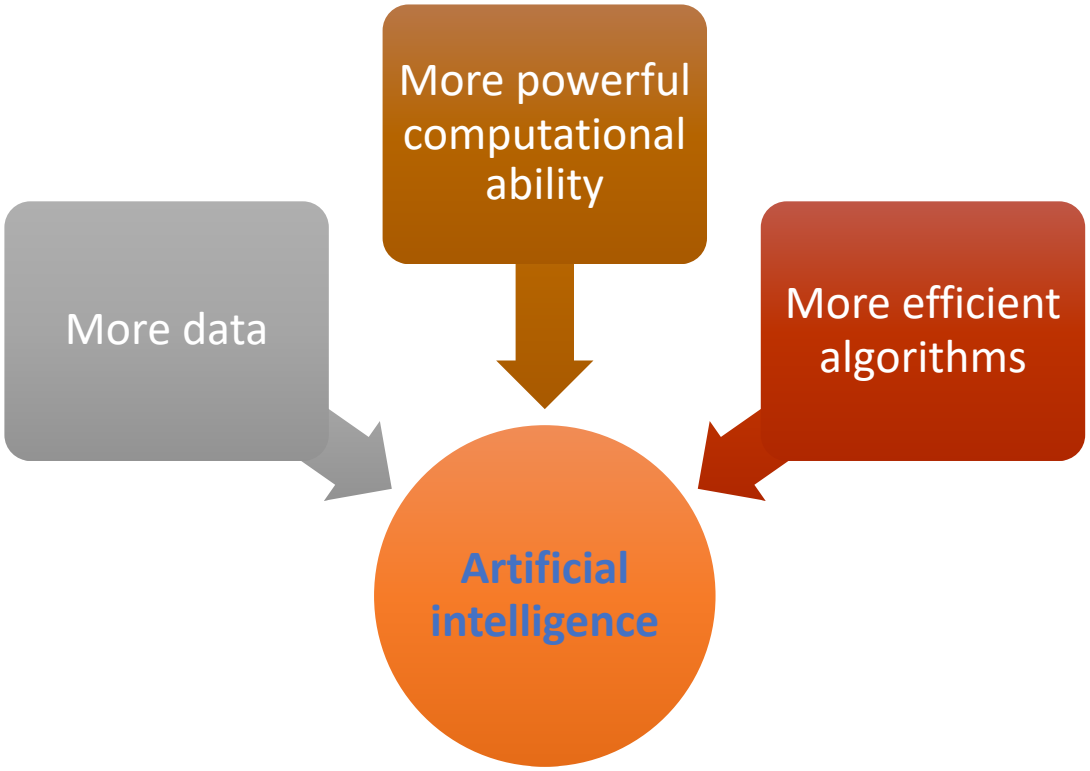
Data science

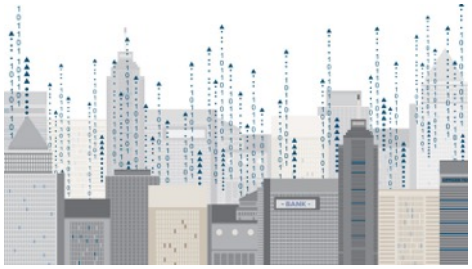
Example: different approaches to language learning

- Non-Native speakers
  - Rules of pronunciation, syntax, …
- Native speakers
  - Imitating ….

Which one is more scientific or more effective?

The era of the artificial intelligence





Can a machine tell the difference:

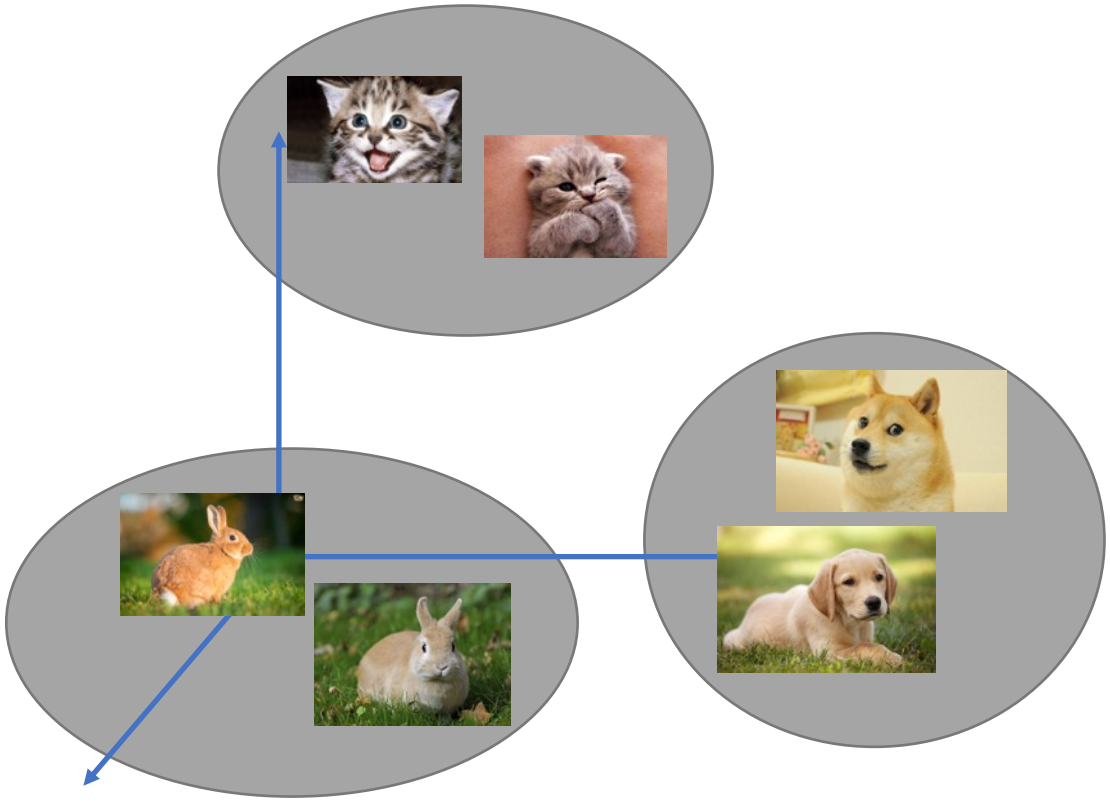


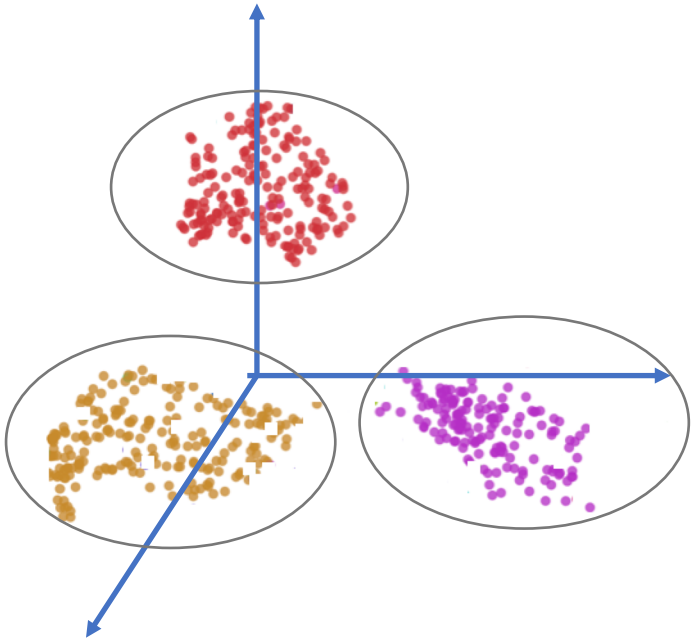



each image = a big vector of pixel values

Size N=1280*720*3 (width*height*RGB channel) ≈ 3M

- 3 different sets of points in R^N, are they separable?





One core problem of AI: clustering

One simple example

- A collection of points R^N is separated into several classes somehow …
- The object is to construct a function y=f(x). The input x of the function is the coordinated of the points in the space and the output y is the cluster the point belongs to.

Graph Recognition

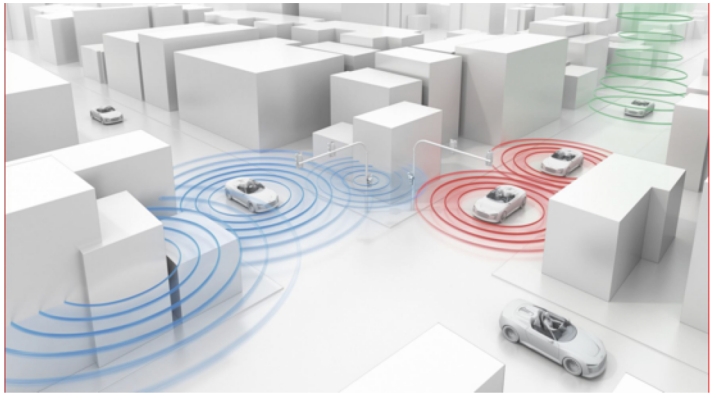
- In 1998, LeCun etc. proposed the neural network LeNet-5 based on the convolution and applied to the hand-written digits recognition successfully. Hence **LeNet-5** is also called the first convolutional neural network(CNN) successfully applied.
- In 2012, Hinton and his student Alex joined the graph recognition competition by ImageNet and improved the accuracy significantly with the CNN **AlexNet**.
- In 2013 Google purchased a Canadian startup company on neural network, **DNNResearch**. It's set up by Geoffrey Hinton and his graduate students Alex Krizhevsky and Ilya Sutskever in 2012.
- In 2015, Kaiming He etc. proposed the **ResNet** CNN structure, which has become a state-of-art CNN structure and widely used in industry and discussed in academia.
- Afterwards CNN is widely used in the field of computer vision and breaks the records ceaselessly.

AlphaGo & AlphaZero

- From 2016 to 2017, the AI program AlphaGo developed by Google DeepMind beat down all the Go champions worldwide.
- 2018 AlphaZero gave a unified principle for many other board games.
- The CEO of Google DeepMind Demis Hassabis announced to integrate AlphaGo with medical, robots and so on. They can learn by themselves since they are artificial intelligence, and transfer learning can be done with enough data.

Auto-driving (the next competition in AI)

- In 2009, Google proposed the plan to replace human driving with softwares.
- Afterwards, many large technology companies such as Tesla, Google, Uber and Benz devoted a lot on investigating the technology of automated driving.
- In many countries, road examination autonomous vehicles are allowed with applications.
- In China, Baidu established the division of intelligent car and opened the platform of automated driving to the public.
- In China, there are about dozens of companies focus on automatic drive techniques from L2 to L4.



Speech recognition

- Recently, the application of recurrent neural network(RNN) and attention model on the speech recognition has achieved great breakthrough on the accuracy of AI speech recognition.
- On October 18 2016, during the Smartisan mobile phone launch event, the input speed of **iFLY** surpassed that of the real keyboards and achieved 100% in accuracy.
- In Nov 2016, IFLYTEK, Baidu, Sogou announced they achieved 97% in the accuracy of Chinese speech recognition respectively.
- 2018, Google propose the **BERT** as a evolutionary model in NLP field.

Machine translation

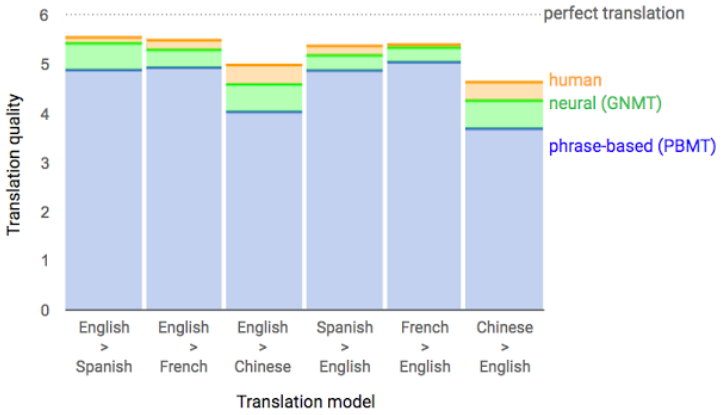
- On Sep 27 2016, Google posed a paper in ArXiv.org about Google's Neural Machine Translation (GNMT). On 28th, to the extremely difficult Chinese-English translation.

- 2018, Baidu, Microsoft, Google announced that they proposed some model to **translate better than humans** in some certain fields and languages.

Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi
yonghui,schuster,zhifengc,qvl,mnorouzi@google.com

Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Łukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, Jeffrey Dean

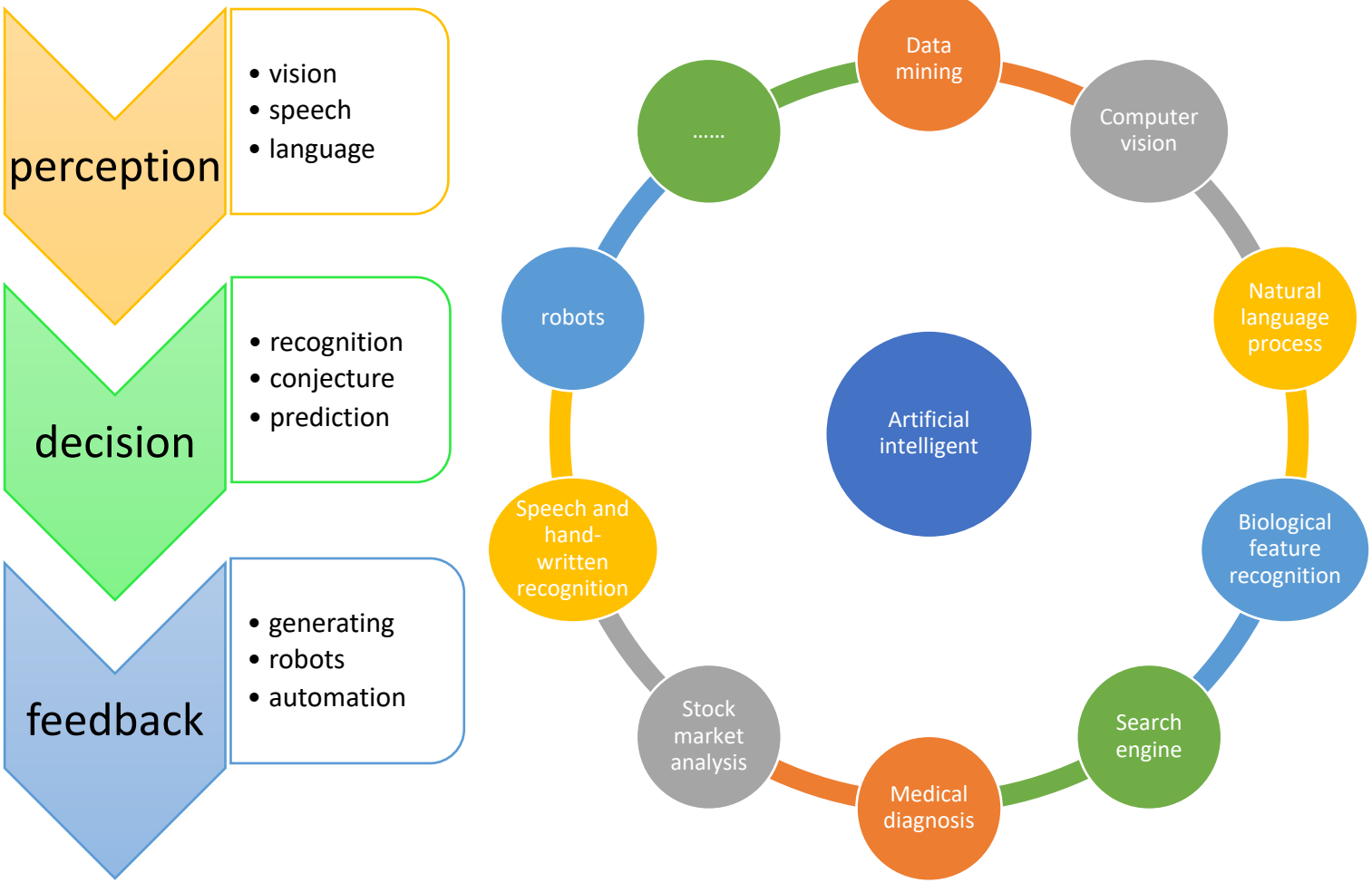


Future smart healthcare

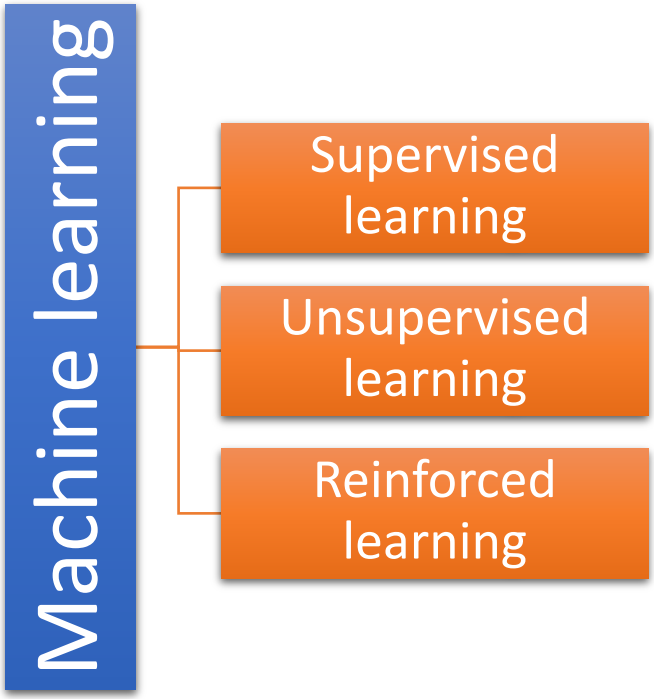
- In 2015, Watson diagnosed leukemia for a 60-year old woman within 10 minutes and proposed adequate therapeutic schedule to Institute of medical science, University of Tokyo
- On the world cancer day in 2017, Watson wrote a prescription for a patient with advanced gastric cancer within 10 seconds in Tianjin third central hospital
- According to the news of Harvard University, recently some softwares of tumor diagnosis developed by deep learning have beat the expert doctors with rich experience in accuracy
- The future scope of the big data on medical and smart healthcare is broad.

IBM Watson

Artificial intelligent



Machine learning

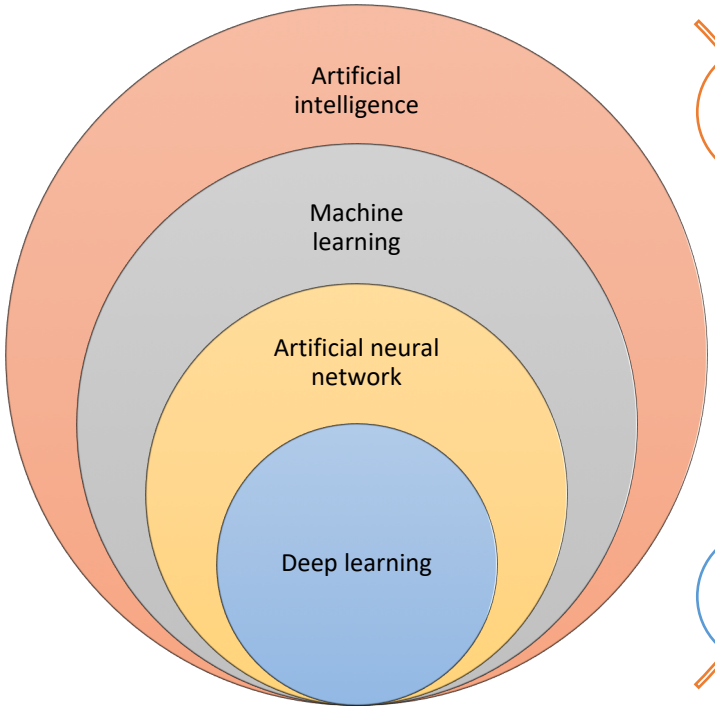


Learn a pattern from a given training datasets and make prediction with this pattern for new data.

Learn the information in the data with samples without labels, such as data clustering, dimension reduction and so on.

Reinforced learning is a field of machine learning, focusing on actions based on the environment to maximize the expected gains.

Deep learning



The object of the artificial intelligence is to have the machines to model the human intelligence

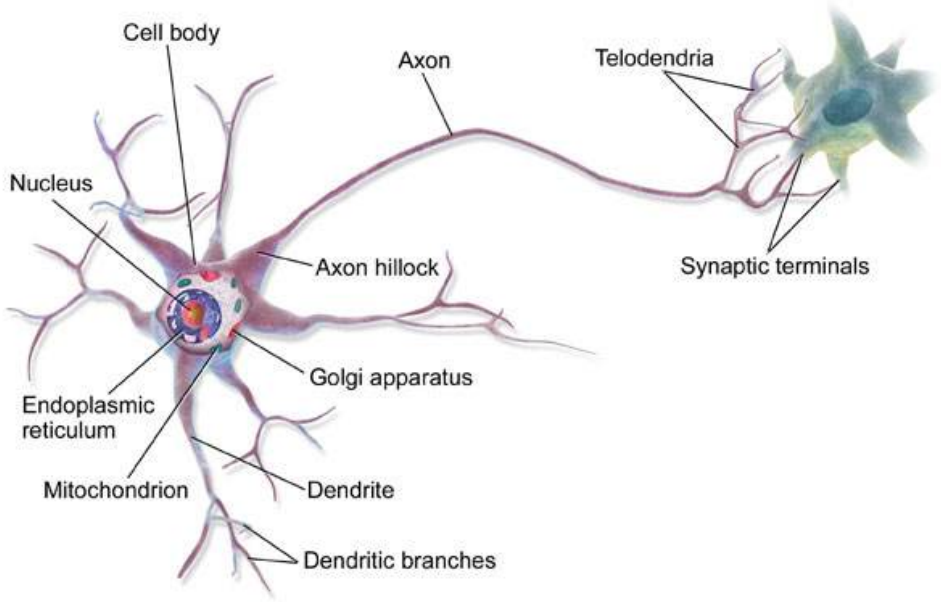
The core of the AI is the machine learning, which is the fundamental way to equip the computers with intelligence

Artificial neural network is an algorithm in early machine learning, consisting of perception, multilayer perception and deep learning

The motivation of the deep learning is to establish and model the human brain in the process of analyzing and learning

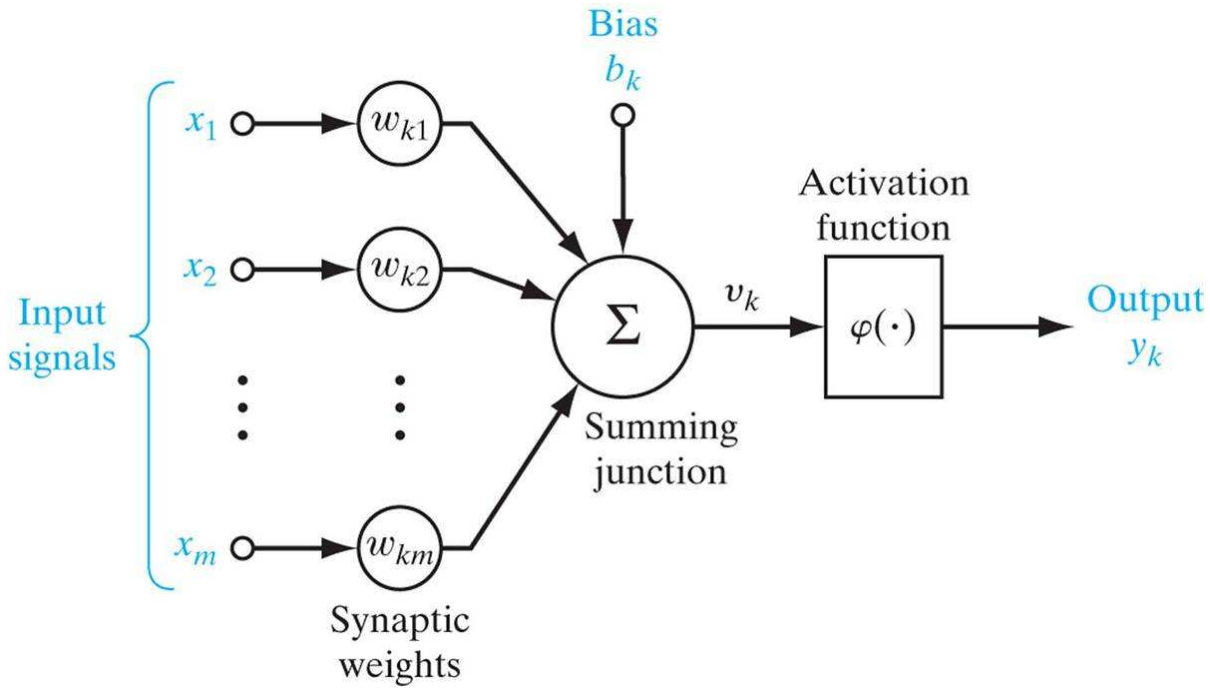
Neurons of animals

- Dendrite: receive the electric signal from other neurons
- Cell body: process the electric signal from other neurons
- Axon: send the electric signal to other neurons
- Activation: the intensity of the signal is larger than the threshold



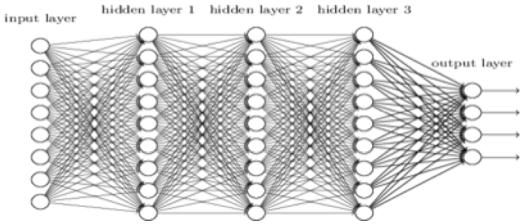
Artificial neuron

- An artificial neuron consists of
  - An affine transformation: $g(x) = Wx + b$
  - A nonlinear function(activation fuction): φ
- The output of the neuron is $y = \varphi(g(x))$



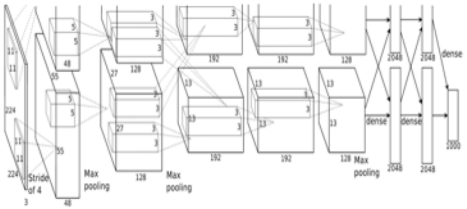
Major models in deep learning

Deep neural network



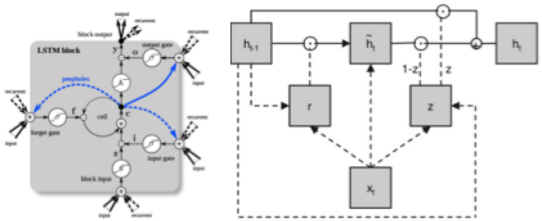
- Model with nonlinear statistical data
- Data classification
- Feature extraction

Convolutional neural network



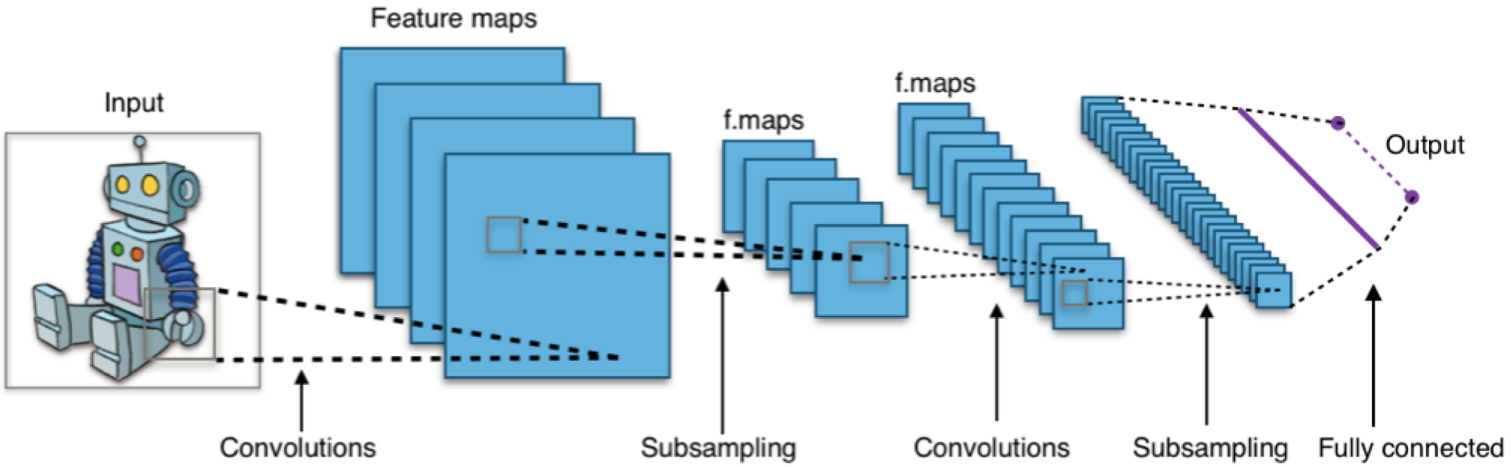
- Graph recognition
- Graph segmentation
- Video analysis
- Natural language processing

Recurrent/recursion neural network



- Speech recognition
- Natural language processing
- Intelligent dialogue
- Machine translation

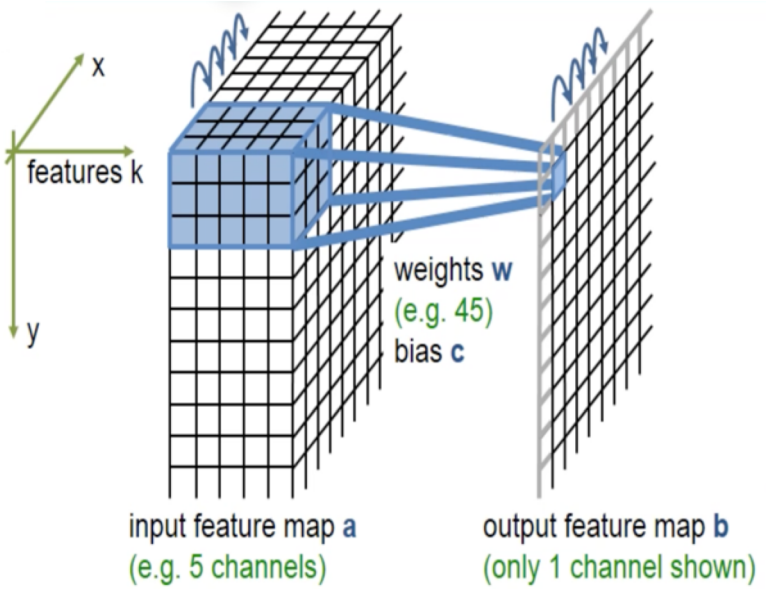
Convolutional neural network(CNN)

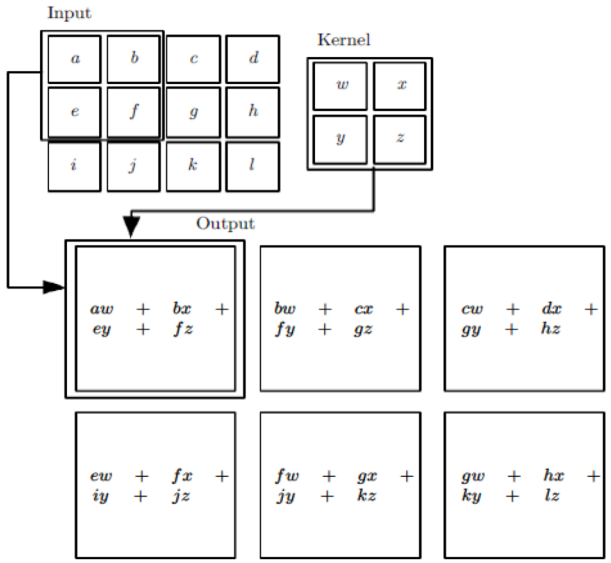


Convolutional neural network consists of

- Convolution layer
- Nonlinear layer(activation function)
- Pooling layer
- Fully-connected layer

Convolutional layer

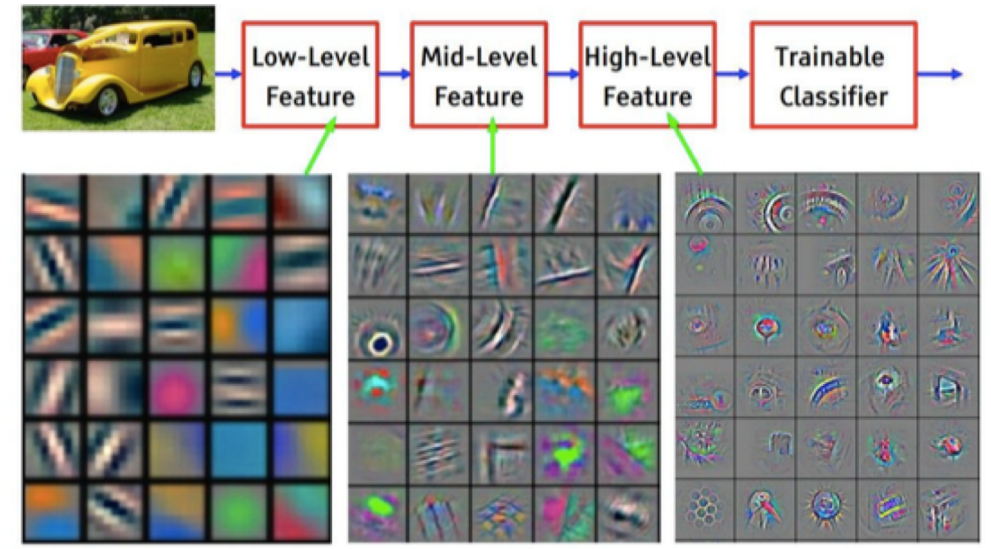




$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) K(m,n)$$

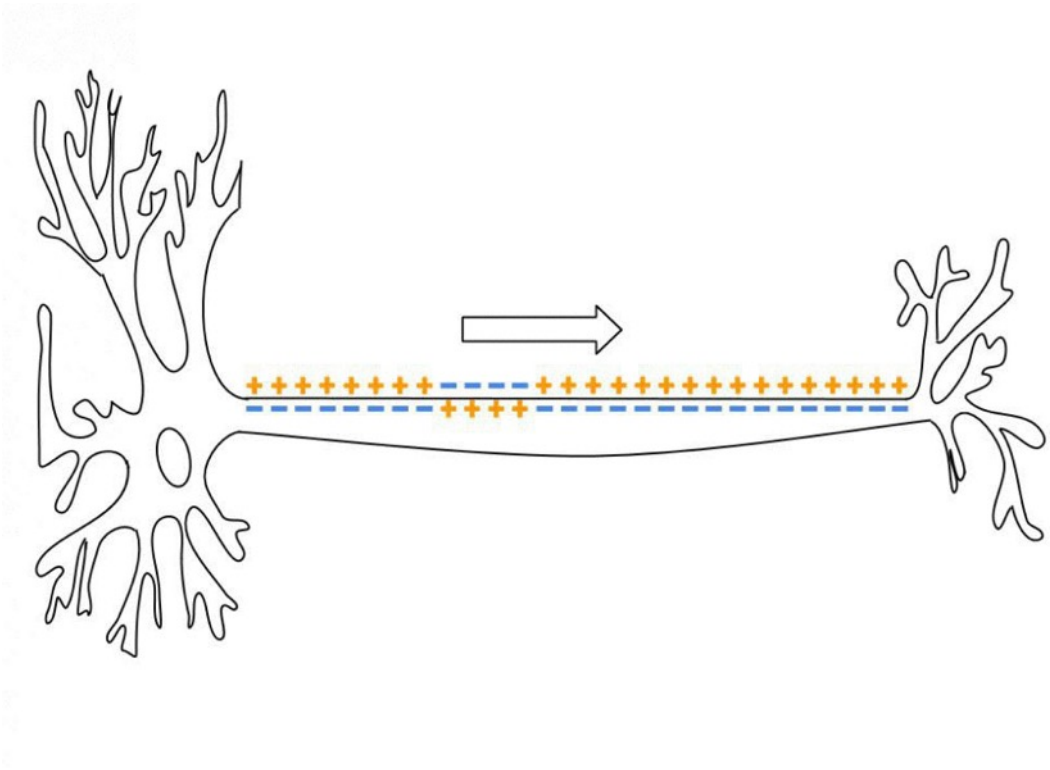
Convolutional kernel

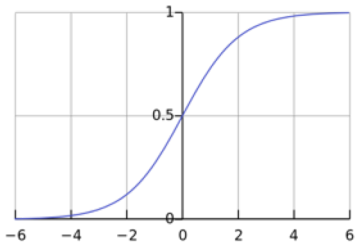
- A convolutional layer consists of multiple convolutional kernel
- Each convolutional kernel extract a certain feature/channel
- The graph processed after convolutional layer is called feature mapping



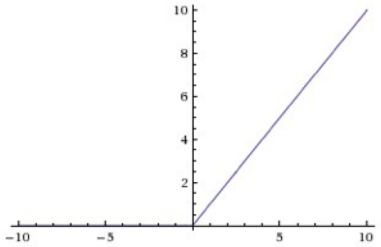
Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Activation function



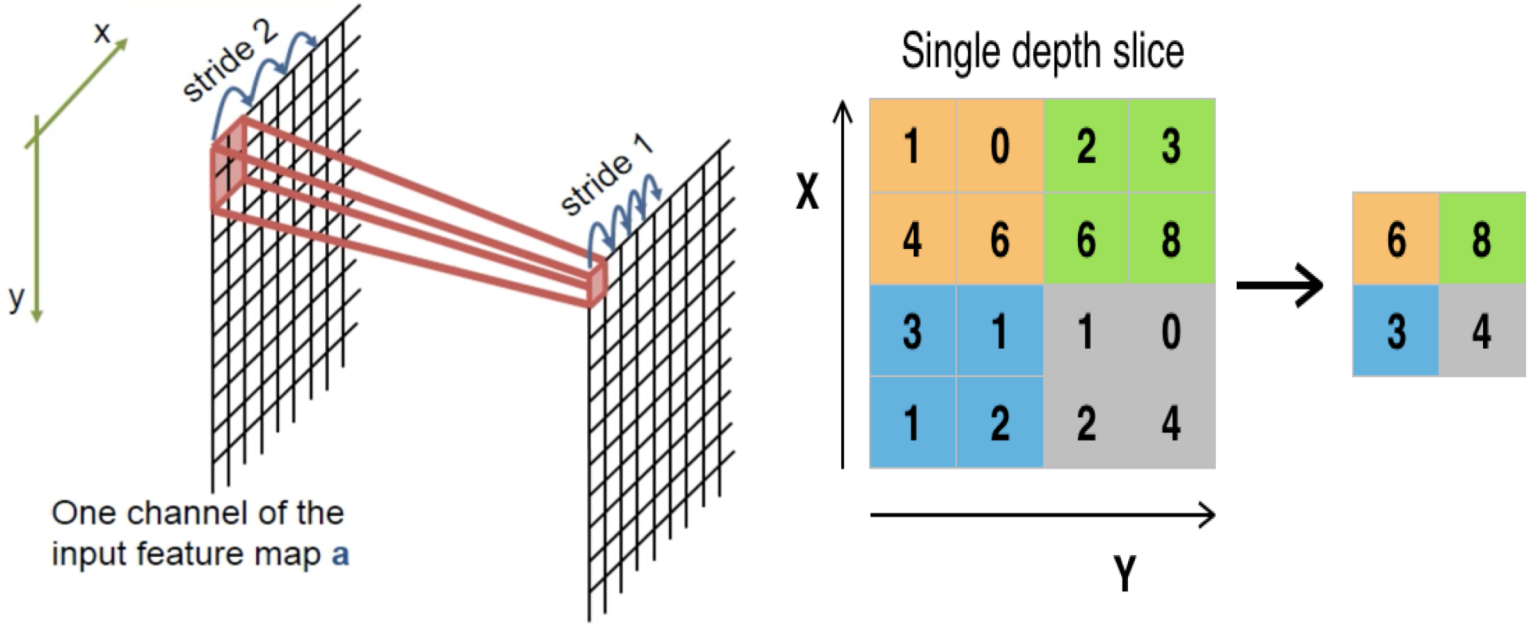


Sigmoid $\sigma(x) = \frac{1}{1+e^{-x}}$

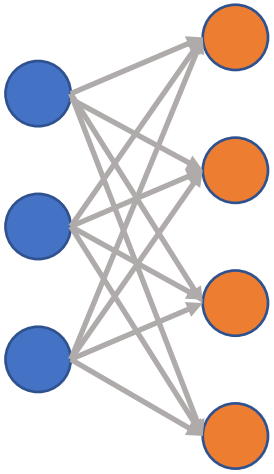


ReLu $\sigma(x) = \max(0, x)$

Pooling layer



Fully-connected layer



- Each neuron in the previous layer is connected with each neuron in the latter layer
- A fully-connected layer can be represented with the following linear function

$$y = Wx+\mathrm{b}$$

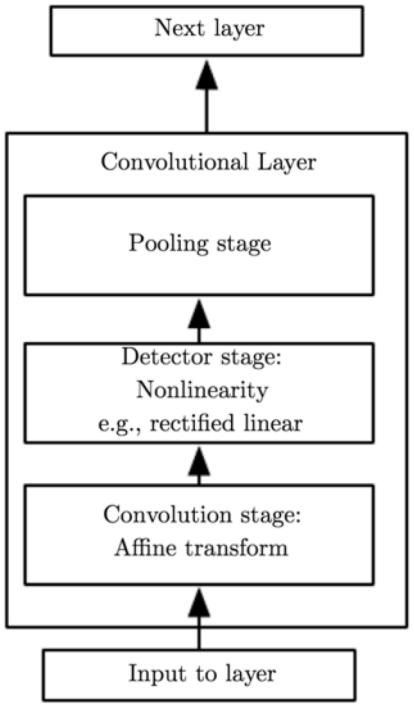
x --- input

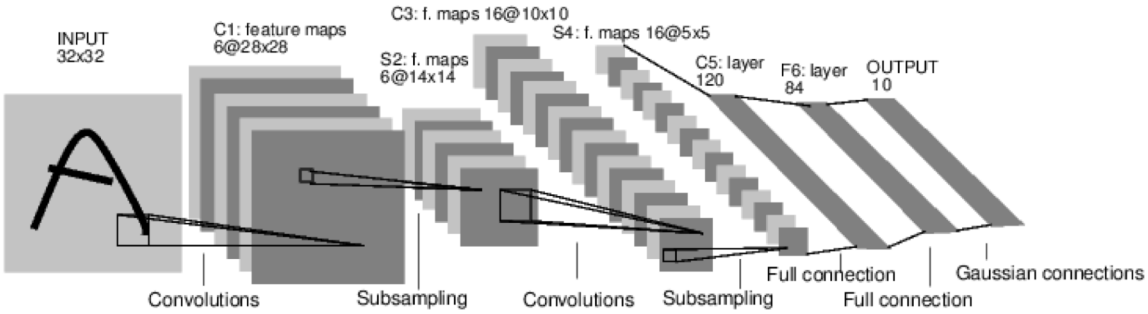
y --- output

W, b--- parameters

- Fully-connected layer makes classification with the features extracted by the convolutional layer

A complete convolutional neural network

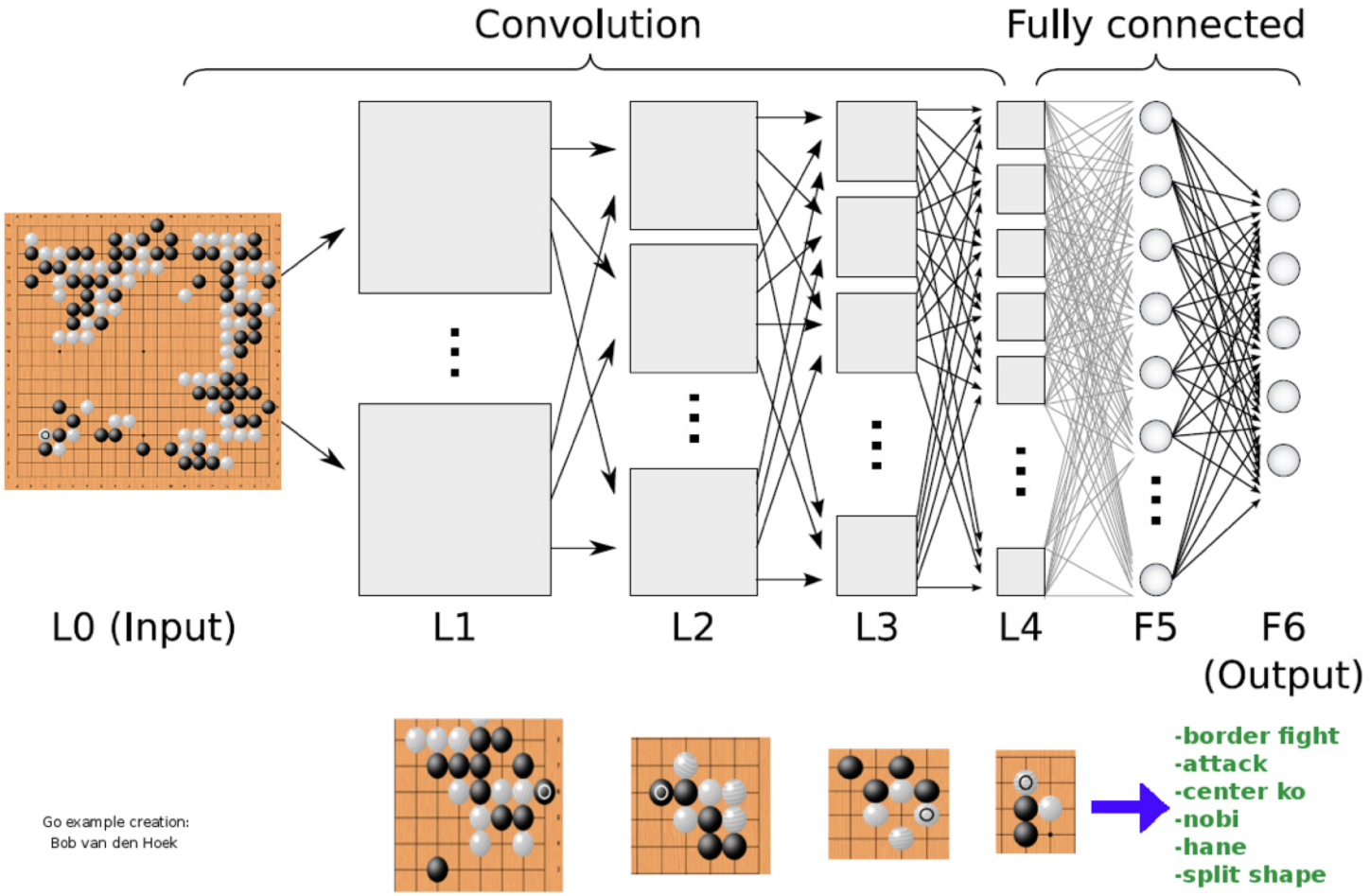




First convolutional neural network: LeNet-5 (Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, 1998)

In the convolutional neural network, convolutional layer is used to extract the features of the graphs and the fully-connected layer is used to make classification with the extracted features.

An example of convolutional neural network: AlphaGo

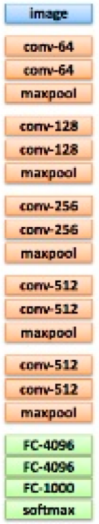


Deep convolutional neural network

8 layers

AlexNet (2012)

19 layers



VGG(2014)

22 layers

GoogleNet(2014)

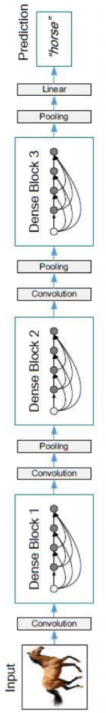
152 layers

ResNet(2015)

1001 layers

iResNet(2016)

160 layers



DenseNet(2017)

Learning(training)

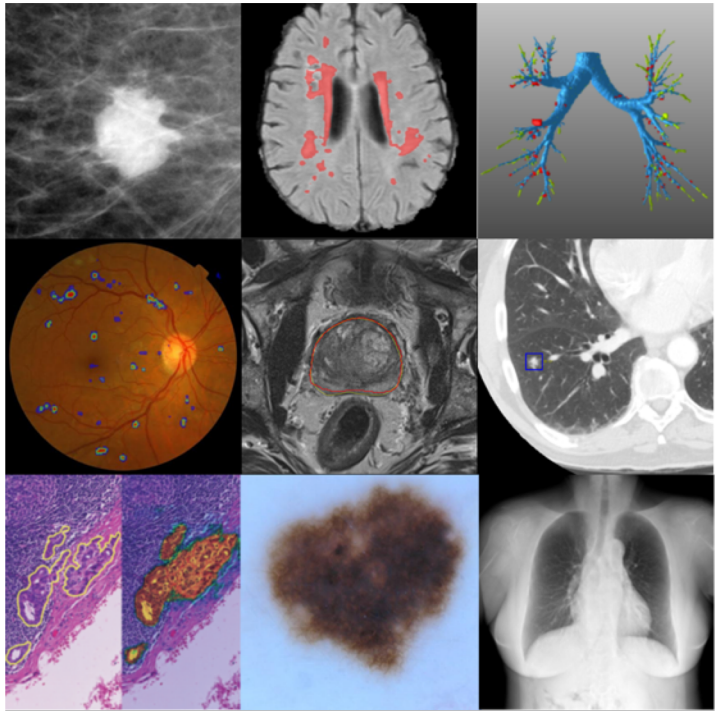
The process of learning is the process to minimize the loss function

$$\min_{W \in \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^{m} \mathcal{L}(f(X^i; W), Y^i)$$

- Training methods in common use: gradient descend, stochastic gradient descend, simulating annealing, etc.
- The training time is extremely long for the current deep neural network due to the plentiful parameters. It takes several weeks to train a traditional deep learning model
  - AlexNet: takes 5-6 days to train on two GTX 580 3GB GPU
  - VGGNet: takes 2-3 weeks to train on NVIDIA Titan Black GPU
  - ResNet-50: 30 hours on DGX-1, 224 seconds on (Nvidia V100*2176 ≈ $13M)
- Research direction：improvement on the models and training algorithms
  - Construct more efficient neural network model: deep or shallow?
  - Implement efficient and stable training algorithms
  - GPU? TPU? Or CPU only?

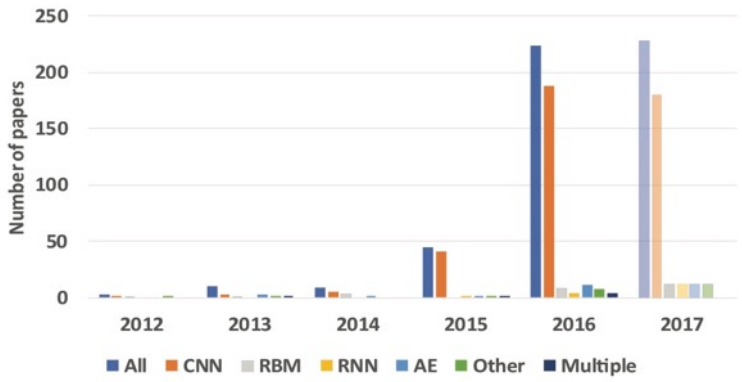
AI + Medical

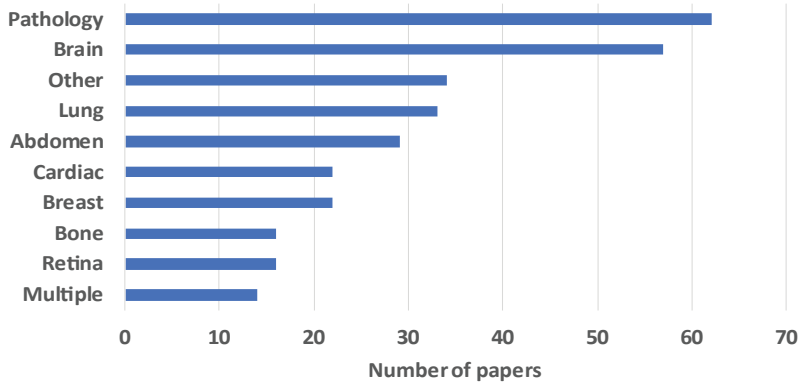
- Images
  - Medical images screening
  - Image registration
  - Organ segmentation in images
  - Diseased region extraction
  - Lesion feature classification
- Biological parameters
  - Disease diagnosis or alert based on the blood glucose, heart rate, ECG, etc.
  - Noninvasive detection of biological parameters such as blood glucose and blood fat
- others
  - Operation result prediction
  - Online medical



Medical AI

- 1970s --- 1990s: traditional methods: manual set rules + mathematical modeling
- 1990s --- 2000s: manual feature extraction+ supervised learning
- 2000s --- 2010s: end to end(automatic feature extraction+ learning)



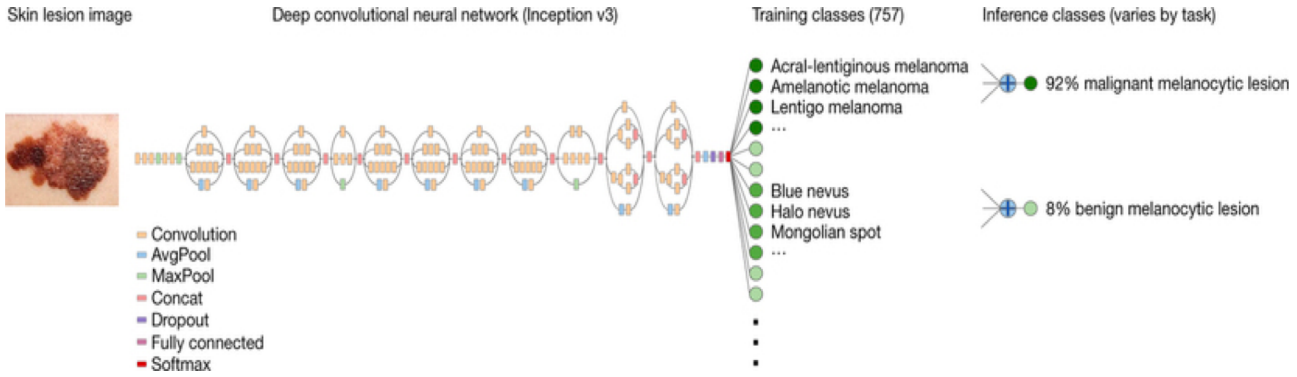


The number of papers about the AI in the medical field (Litjens et.al, 2017)

Medical AI- I

AI can automatically diagnose the cancer

- In 2017, a team in Stanford University achieved AI automatic diagnosis of the skin cancer with the CNN, the accuracy of which is as high as a human expert.
- This model is trained based on a public model by Google, while the original model is only used to classify the cats and dogs in photos.

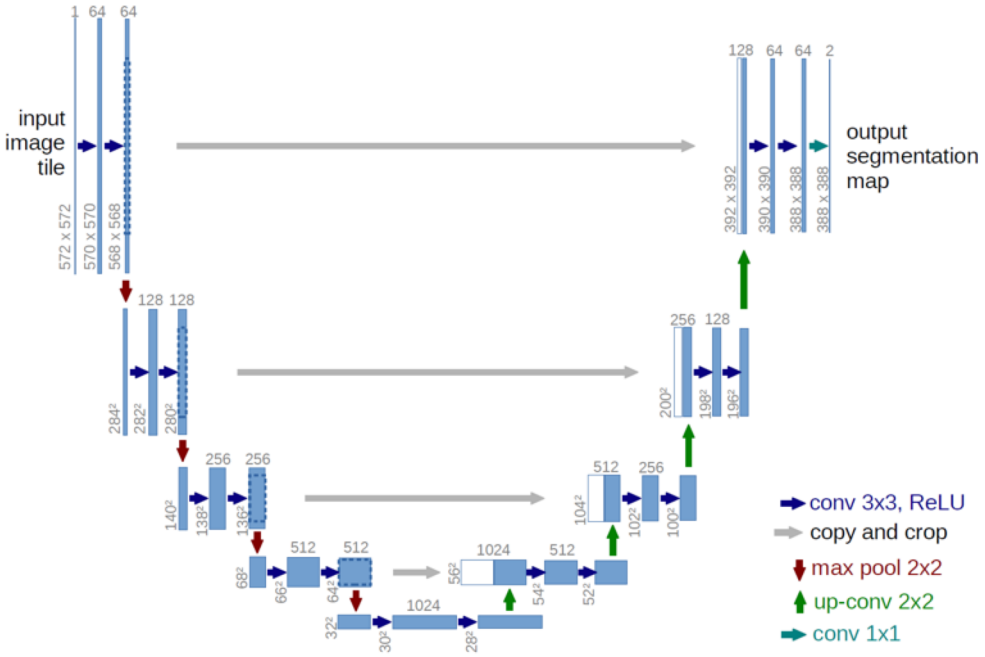


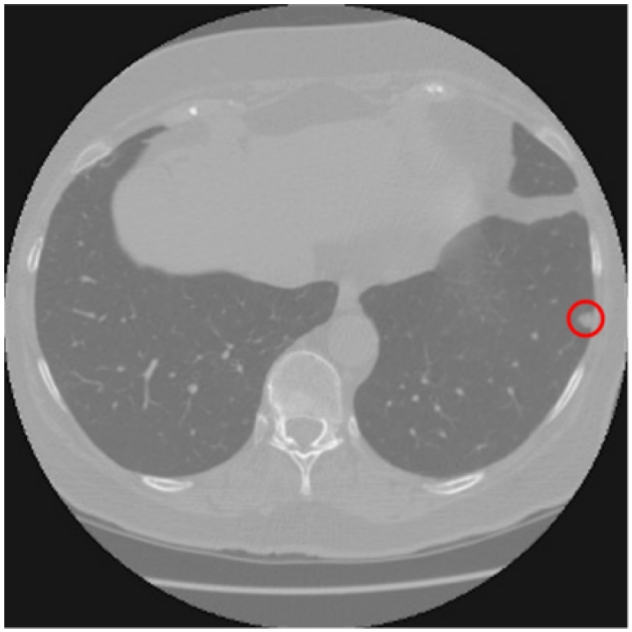
Ref: *Dermatologist-level classification of skin cancer with deep neural networks*, Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun, *Nature* 542.7639 (2017): 115-118.

Medical AI- II

The deep neural network has achieved the recognition of focus position in medical images：

- During 2016 to 2017， two famous data competition platforms Kaggle and Ali yuntianchi held competitions of pulmonary nodules intelligent diagnosis based on the CT images
- The winner team used the U-Net model based on the CNN and achieved high accuracy





Medical AI- III

AI can generate the diagnosis report automatically

- By combining CNN and RNN, learned with X-Ray images of chest and corresponding diagnosis report data, it can be achieved for classification of chest X-Ray and generate the diagnosis report automatically

Two views, XXXX and lateral

Comparison: None available

Indication: XXXX-year-old male for preop XXXX of subclavian arterial stent.

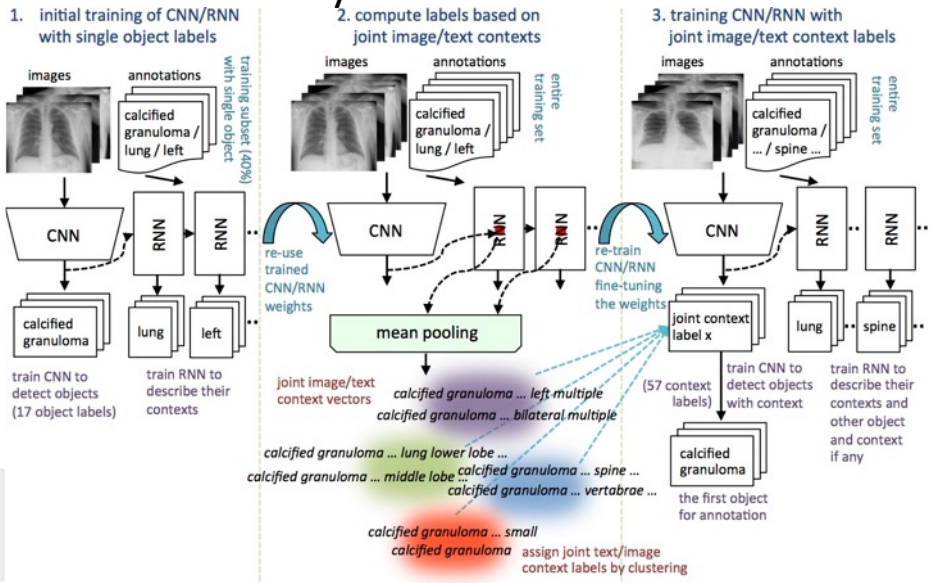
Findings: Cardiac and mediastinal contours are unremarkable. Pulmonary vascularity is within normal limits. No focal air space opacities, pleural effusion, or pneumothorax. There are increased lucencies in the bilateral apices along with horizontal oblique scarring in the left upper lobe. This could suggest emphysematous bullae. XXXX are grossly unremarkable.

Impression: No active disease.

MeSH

Major

Lung, Hyperlucent / apex / bilateral
Lucency / lung / apex / bilateral /
Cicatrix / lung / upper lobe / left

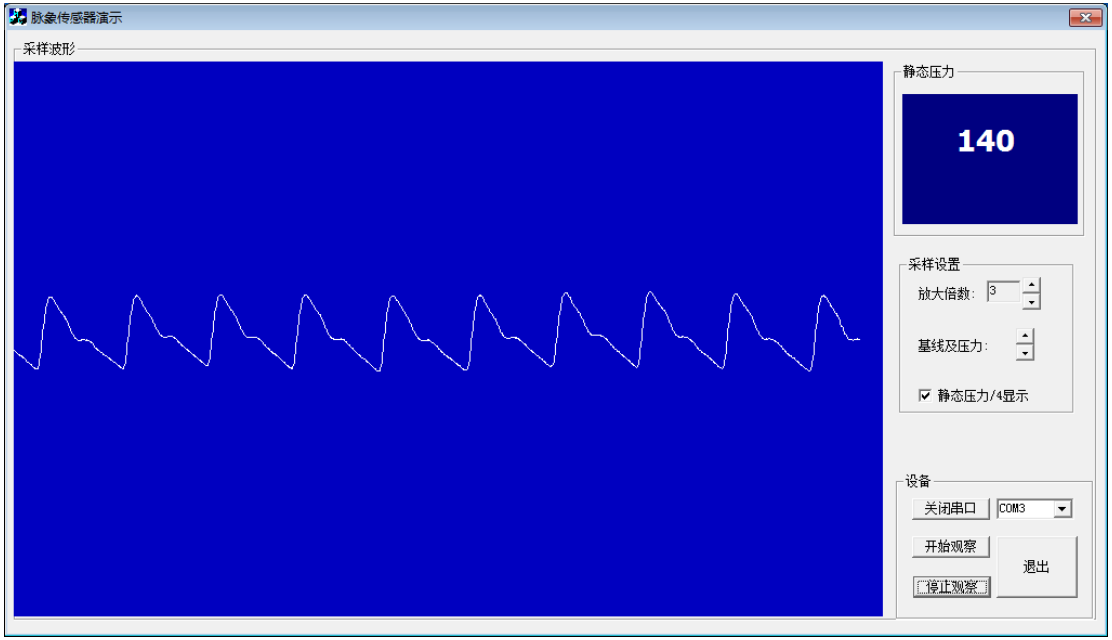


Ref: *Learning to Read Chest X-Rays: Recurrent Neural Cascade Model for Automated Image Annotation*, Hoo-Chang Shin, Kirk Roberts, Le Lu, Dina Demner-Fushman, Jianhua Yao, Ronald M Summers, 2016

Medical AI- IV

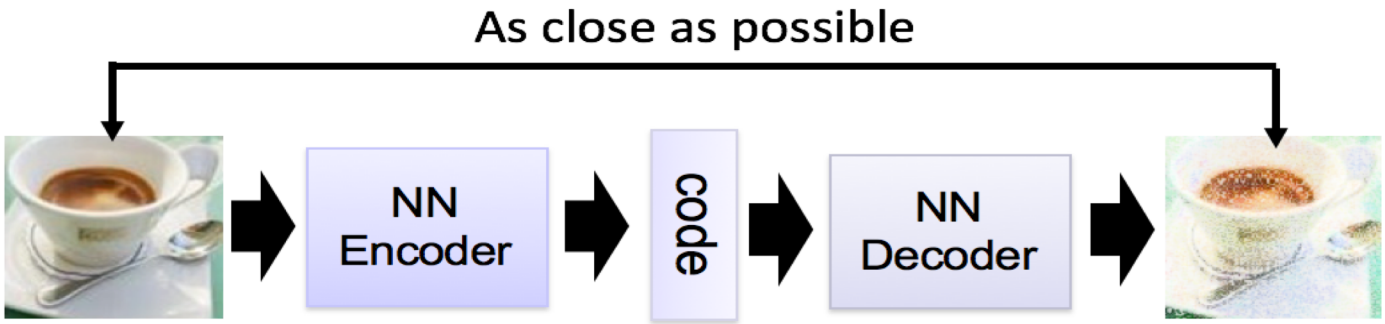
Predict the pregnancy using the pulse wave by convolutional neural n

- The design of the model： Extract the features from the pulse wave by CNN and make classifications
- Current result: Over 85% prediction accuracy

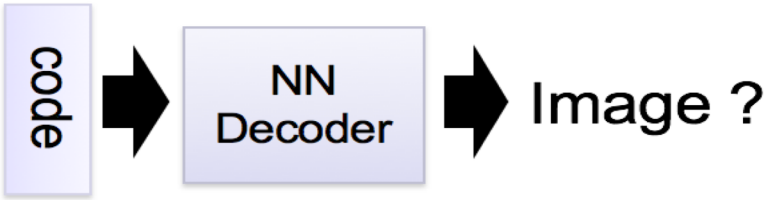


The pulse wave of a pregnant women

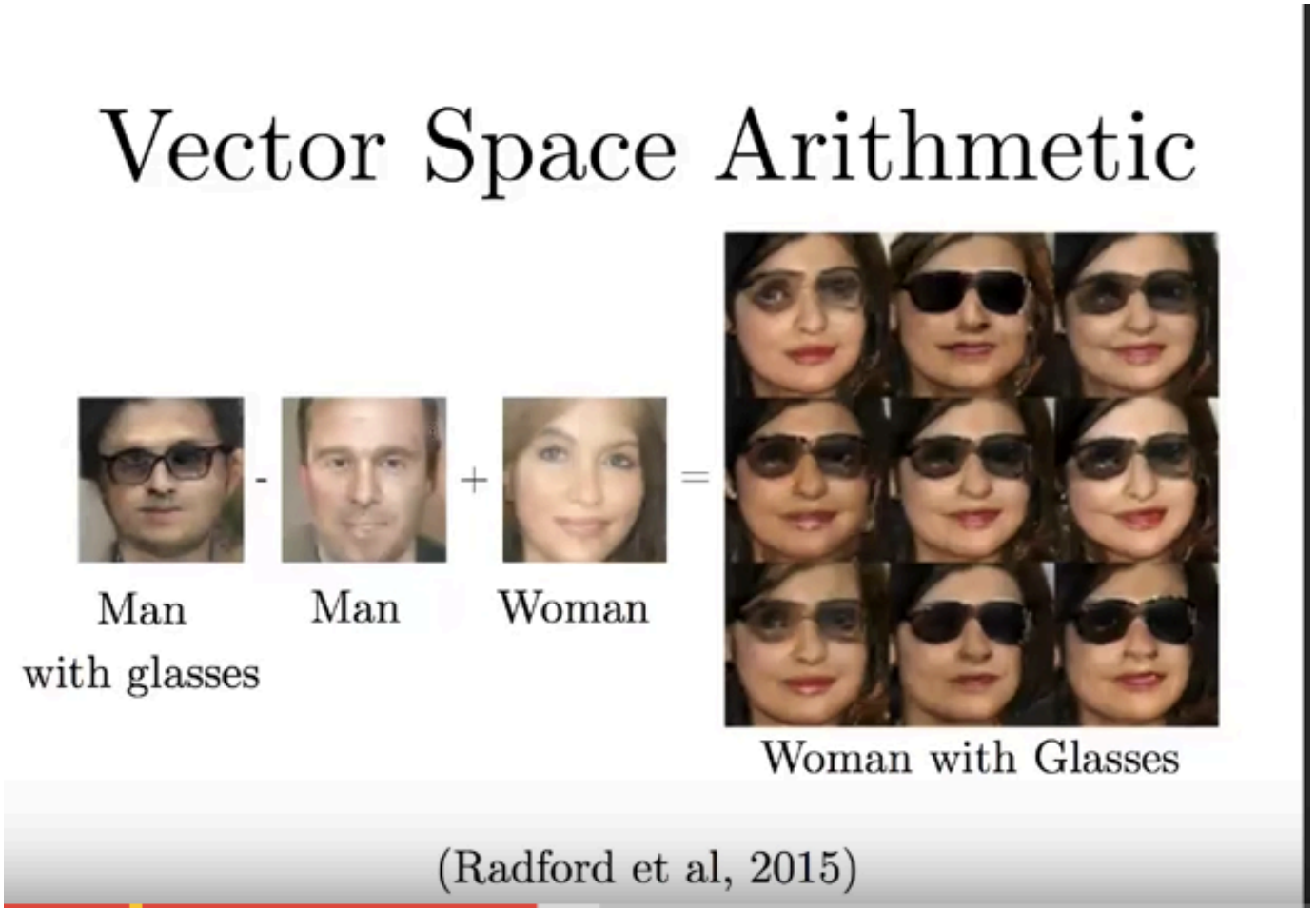
Generative Adversarial Networks



Randomly generate a vector as code



GAN: similar to word embedding



Text to Image, by conditional GAN

Caption	Image
a pitcher is about to throw the ball to the batter	
a group of people on skis stand in the snow	
a man in a wet suit riding a surfboard on a wave	

Text to Image, by conditional GAN

Caption	Image
this flower has white petals and a yellow stamen	
the center is yellow surrounded by wavy dark purple petals	
this flower has lots of small round pink petals