Deep Learning

Dr. Mehran Safayani

safayani@iut.ac.ir
safayani.iut.ac.ir

https://www.aparat.com/mehran.safayani

https://github.com/safayani/deep_learning_course

The slides are modified, based on original slides by [Geoffrey Hinton, Toronto university], [Andrew NG, Stanford university] and [Fei Fei Lee, Justin Johnson and Serena Yeung, standard university].

Department of Electrical and computer engineering, Isfahan university of technology, Isfahan, Iran

Why do we need to neural networks?

What is Machine Learning?

- It is very hard to write programs that solve problems like recognizing a three-dimensional object from a novel viewpoint in new lighting conditions in a cluttered scene.
  - We don't know what program to write because we don't know how its done in our brain.
  - Even if we had a good idea about how to do it, the program might be horrendously complicated.

- It is hard to write a program to compute the probability that a credit card transaction is fraudulent.
  - There may not be any rules that are both simple and reliable. We need to combine a very large number of weak rules.
  - Fraud is a moving target. The program needs to keep changing.

The Machine Learning Approach

- Instead of writing a program by hand for each specific task, we collect lots of examples that specify the correct output for a given input.
- A machine learning algorithm then takes these examples and produces a program that does the job.
  - The program produced by the learning algorithm may look very different from a typical hand-written program. It may contain millions of numbers.
  - If we do it right, the program works for new cases as well as the ones we trained it on.
  - If the data changes the program can change too by training on the new data.
- Massive amounts of computation are now cheaper than paying someone to write a task-specific program.

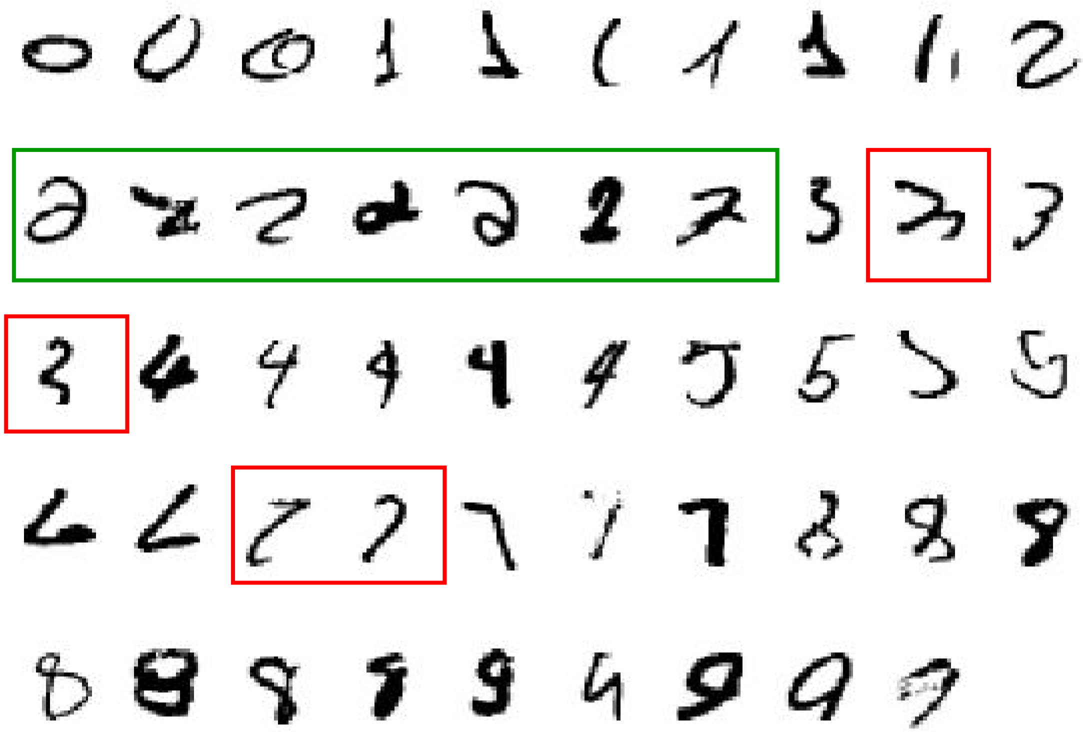
Some examples of tasks best solved by learning

- Recognizing patterns:
  - Objects in real scenes
  - Facial identities or facial expressions
  - Spoken words
- 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?

A standard example of machine learning

- The MNIST database of hand-written digits is the the machine learning equivalent of fruit flies.
  - They are publicly available and we can learn them quite fast in a moderate-sized neural net.
  - We know a huge amount about how well various machine learning methods do on MNIST.
- We will use MNIST as our standard task.

It is very hard to say what makes a 2



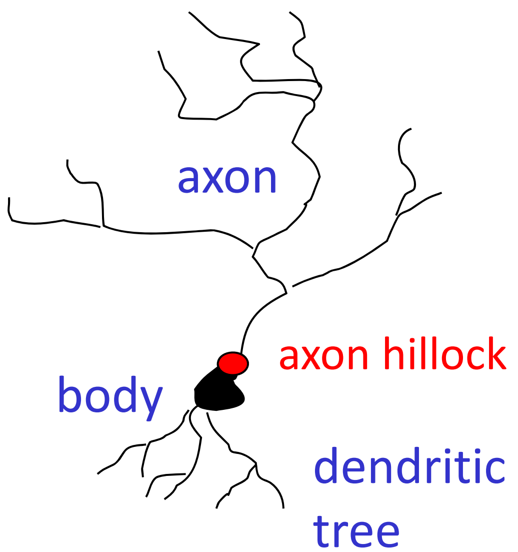
Neural Networks
What are neural networks?

Reasons to study neural computation

- To understand how the brain actually works.
  - Its very big and very complicated and made of stuff that dies when you poke it around. So we need to use computer simulations.
- To understand a style of parallel computation inspired by neurons and their adaptive connections.
  - Very different style from sequential computation.
    - should be good for things that brains are good at (e.g. vision)
- To solve practical problems by using novel learning algorithms inspired by the brain (this course)
  - Learning algorithms can be very useful even if they are not how the brain actually works.

A typical cortical neuron

- Gross physical structure:
  - There is one axon that branches
  - There is a dendritic tree that collects input from other neurons.
- Axons typically contact dendritic trees at synapses
  - A spike of activity in the axon causes charge to be injected into the post-synaptic neuron.
- Spike generation:
  - There is an axon hillock that generates outgoing spikes whenever enough charge has flowed in at synapses to depolarize the cell membrane.

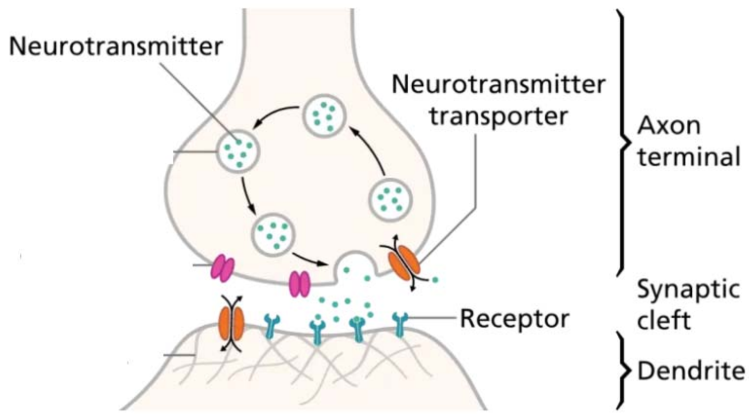


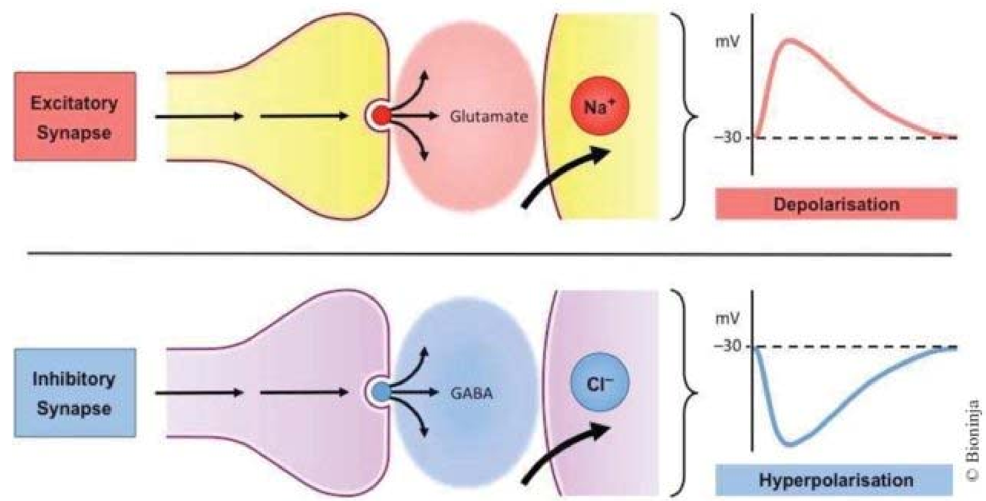
Synapses

- When a spike of activity travels along an axon and arrives at a synapse it causes vesicles of transmitter chemical to be released.
  - There are several kinds of transmitter.

- The transmitter molecules diffuse across the synaptic cleft and bind to receptor molecules in the membrane of the post-synaptic neuron thus changing their shape.
  - This opens up holes that allow specific ions in or out.

Synapses


Neurotransmitter
Neurotransmitter transporter
Axon terminal
Synaptic cleft
Receptor
Dendrite


Excitatory Synapse
Glutamate
Na+
mV
–30
Depolarisation
Inhibitory Synapse
GABA
Cl–
mV
–30
Hyperpolarisation
© Bioninja

How synapses adapt

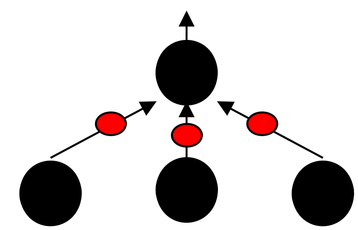
- The effectiveness of the synapse can be changed:
  - vary the number of vesicles of transmitter.
  - vary the number of receptor molecules.

- Synapses are slow, but they have advantages over RAM
  - They are very small and very low-power.
  - They adapt using locally available signals
    - But what rules do they use to decide how to change?

How the brain works on one slide!

- Each neuron receives inputs from other neurons
  - A few neurons also connect to receptors.
  - Cortical neurons use spikes to communicate.
- The effect of each input line on the neuron is controlled

  by a synaptic weight
  - The weights can be positive or negative.
- The synaptic weights adapt so that the whole network learns to perform useful computations
  - Recognizing objects, understanding language, making plans, controlling the body.
- You have about 10^{11} neurons each with about 10^{4} weights.
  - A huge number of weights can affect the computation in a very short time. Much better bandwidth than a workstation.



Modularity and the brain

- Different bits of the cortex do different things.
  - Local damage to the brain has specific effects.
  - Specific tasks increase the blood flow to specific regions.
- But cortex looks pretty much the same all over.
  - Early brain damage makes functions relocate.
- Cortex is made of general purpose stuff that has the ability to turn into special purpose hardware in response to experience.
  - This gives rapid parallel computation plus flexibility.
  - Conventional computers get flexibility by having stored sequential programs, but this requires very fast central processors to perform long sequential computations.

Object Recognition

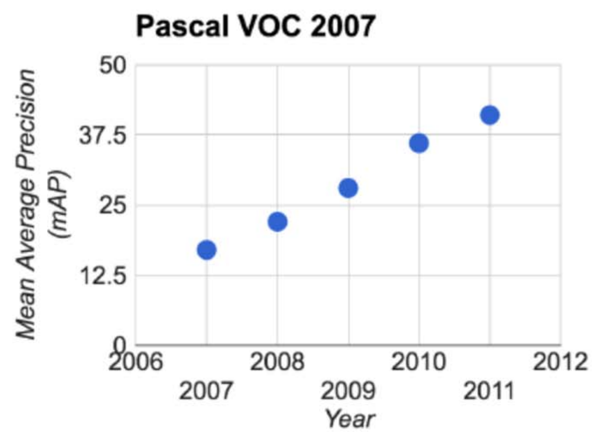
PASCAL Visual Object Challenge
(20 object categories)

[Everingham et al. 2006-2012]

Image is CC0 1.0 public domain

Image is CC0 1.0 public domain

Image is CC0 1.0 public domain



IMAGENET

www.image-net.org

22K categories and **14M** images

- Animals
  - Bird
  - Fish
  - Mammal
  - Invertebrate
- Plants
  - Tree
  - Flower
- Food
- Materials
- Structures
- Artifact
  - Tools
  - Appliances
  - Structures
- Person
- Scenes
  - Indoor
  - Geological Formations
- Sport Activities

IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:
1,000 object classes
1,431,167 images

Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle

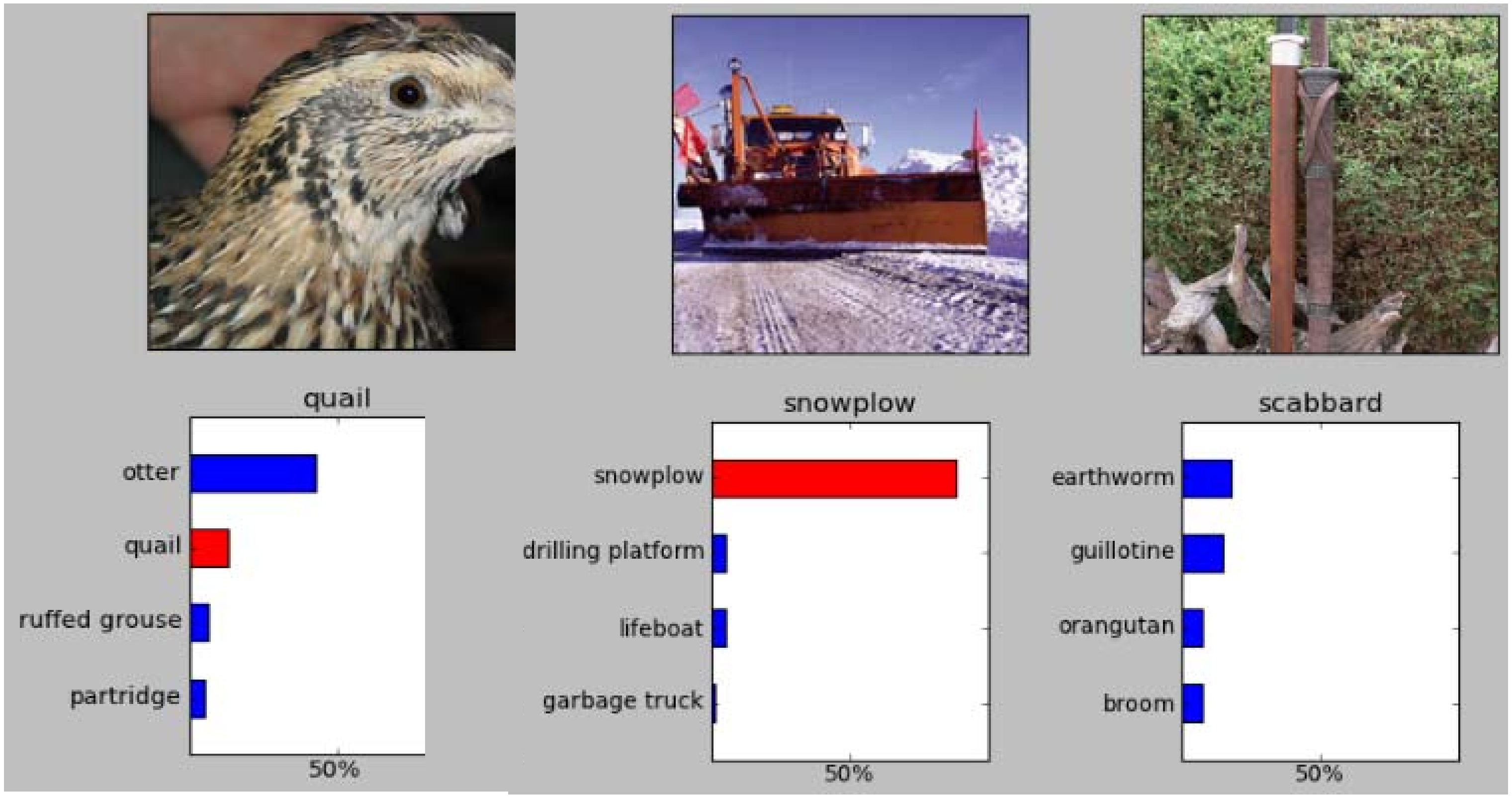
✔

Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle

✗

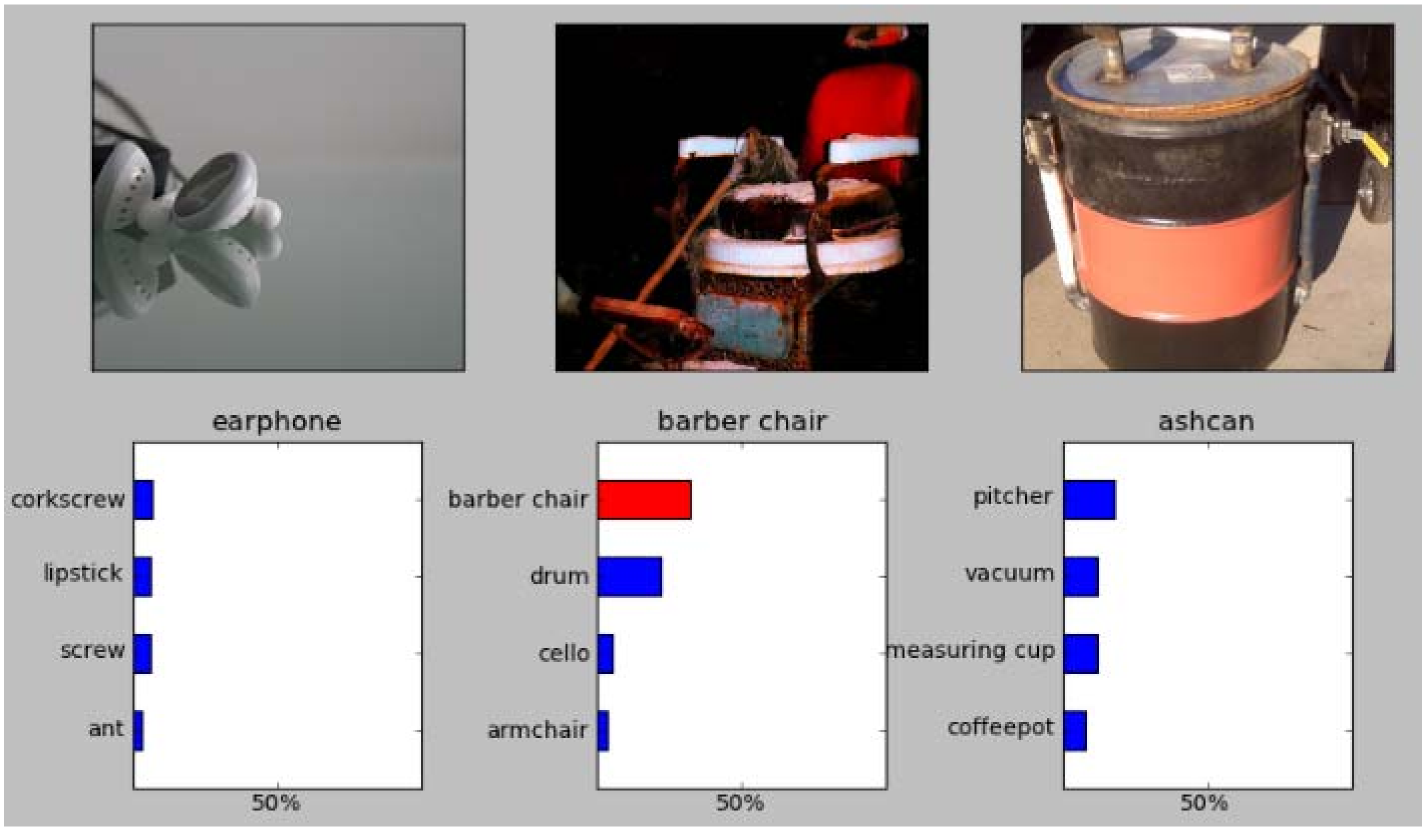
Russakovsky et al. IJCV 2015

Some examples from an earlier version of the net



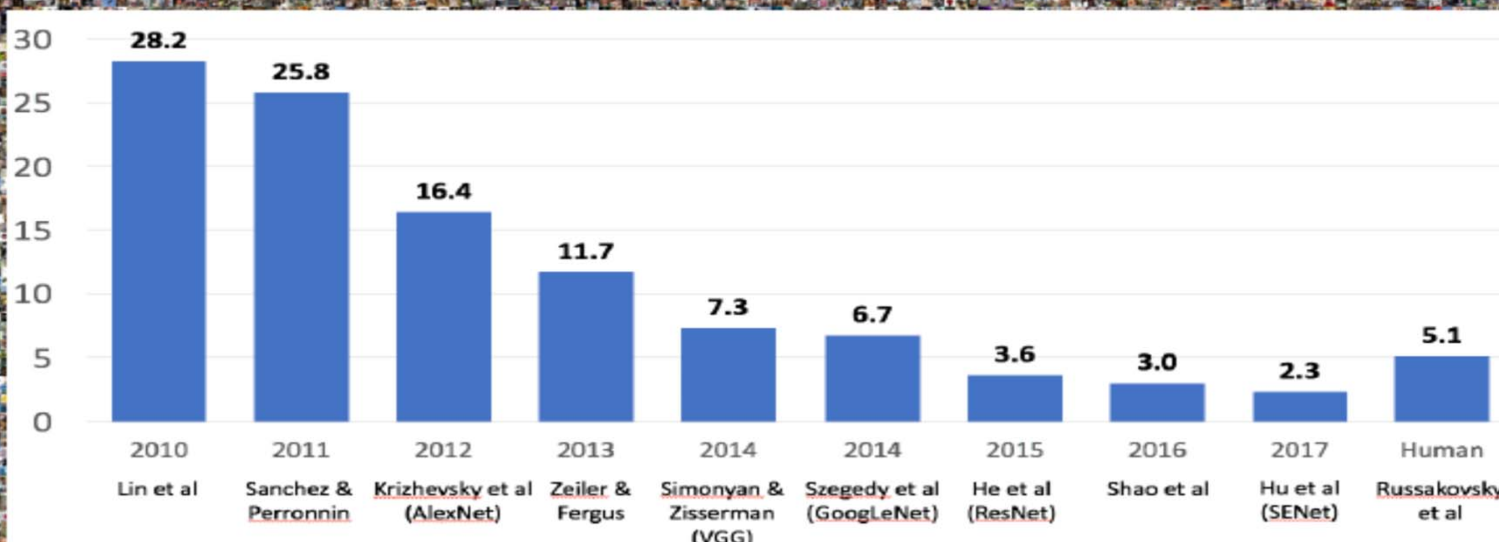
It can deal with a wide range of objects

It makes some really cool errors



IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:
1,000 object classes
1,431,167 images



Russakovsky et al. IJCV 2015

The ImageNet task

- Best system in 2010 competition got 47% error for its first choice and 28% error for its top 5 choices.
- A very deep neural net (Krizhevsky et. al. 2012) gets less that 40% error for its first choice and less than 16% for its top 5 choices
- Hundred million parameters

IMAGENET Large Scale Visual Recognition Challenge

Year 2010

NEC-UIUC

Dense descriptor grid: HOG, LBP

↓

Coding: local coordinate, super-vector

↓

Pooling, SPM

↓

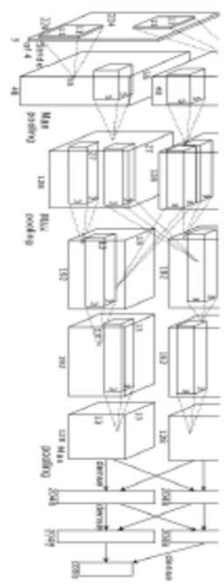
Linear SVM

[Lin CVPR 2011]

Lion image by Swissfrog is licensed under CC BY 3.0

Year 2012

SuperVision

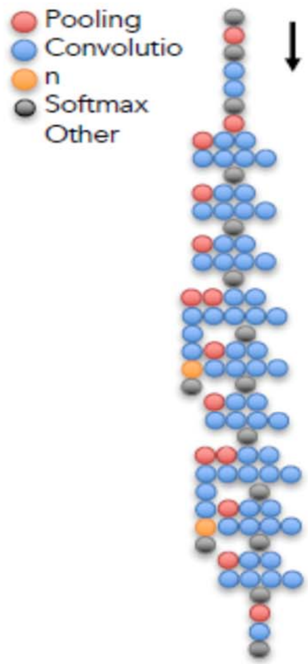


[Krizhevsky NIPS 2012]

Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

Year 2014

GoogLeNet



[Szegedy arxiv 2014]

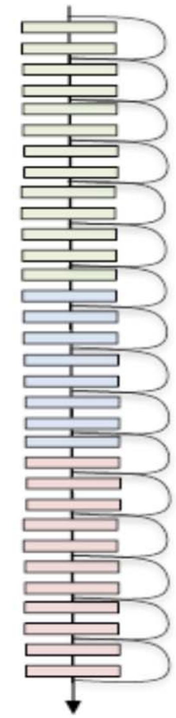
VGG

Image
conv-64
conv-64
maxpool
conv-128
conv-128
maxpool
conv-256
conv-256
maxpool
conv-512
conv-512
maxpool
conv-512
conv-512
maxpool
fc-4096
fc-4096
fc-1000
softmax

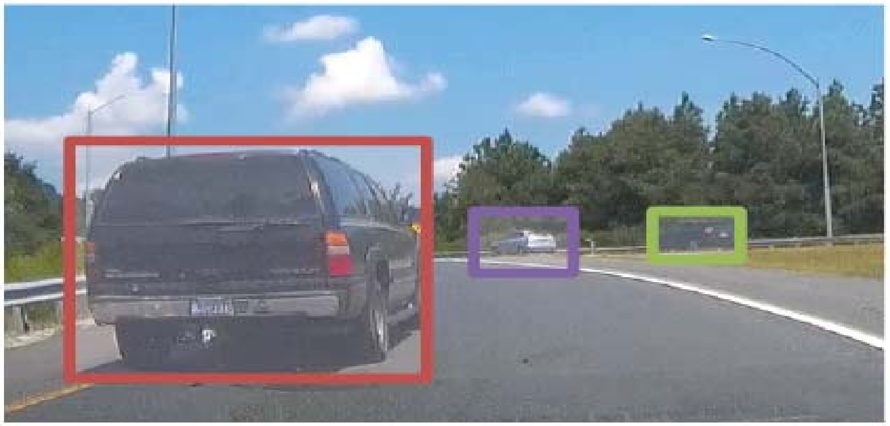
[Simonyan arxiv 2014]

Year 2015

MSRA



[He ICCV 2015]

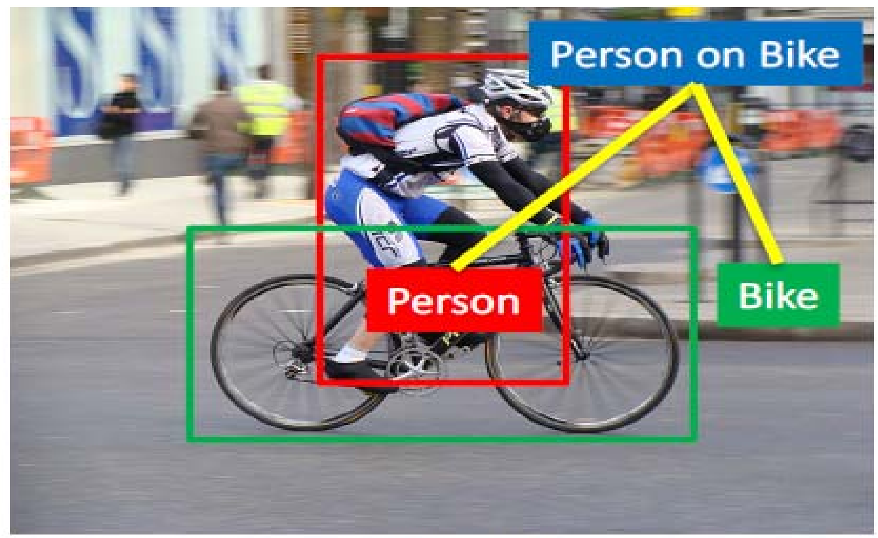


This image is licensed under CC BY-NC-SA 2.0; changes made

- Object detection
- Image captioning

Person

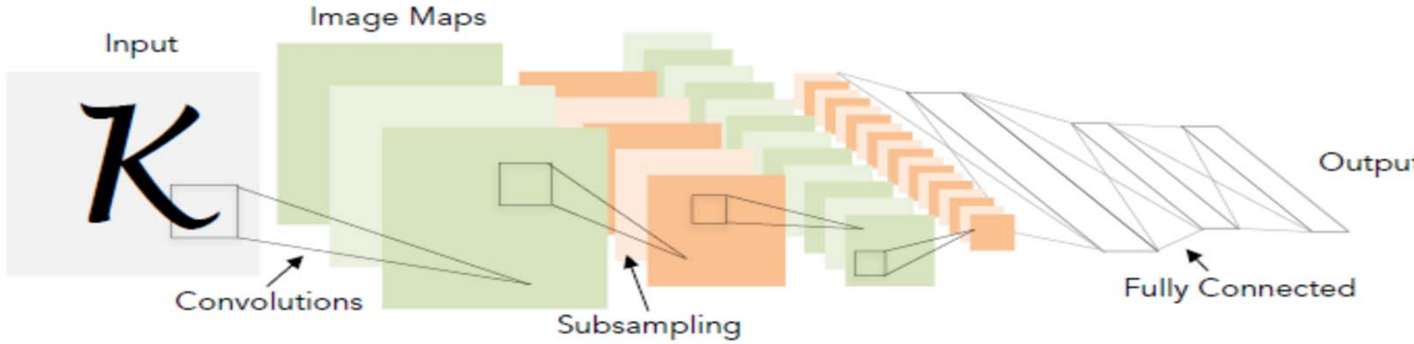
Hammer



This image is licensed under CC BY-SA 3.0; changes made

1998

LeCun et al.



# of transistors

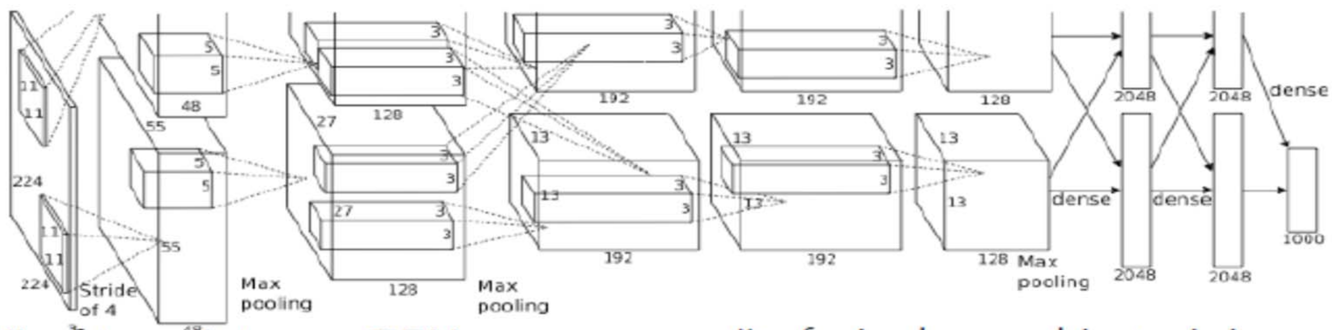
10^6

# of pixels used in training

10^7 NIST

2012

Krizhevsky et al.



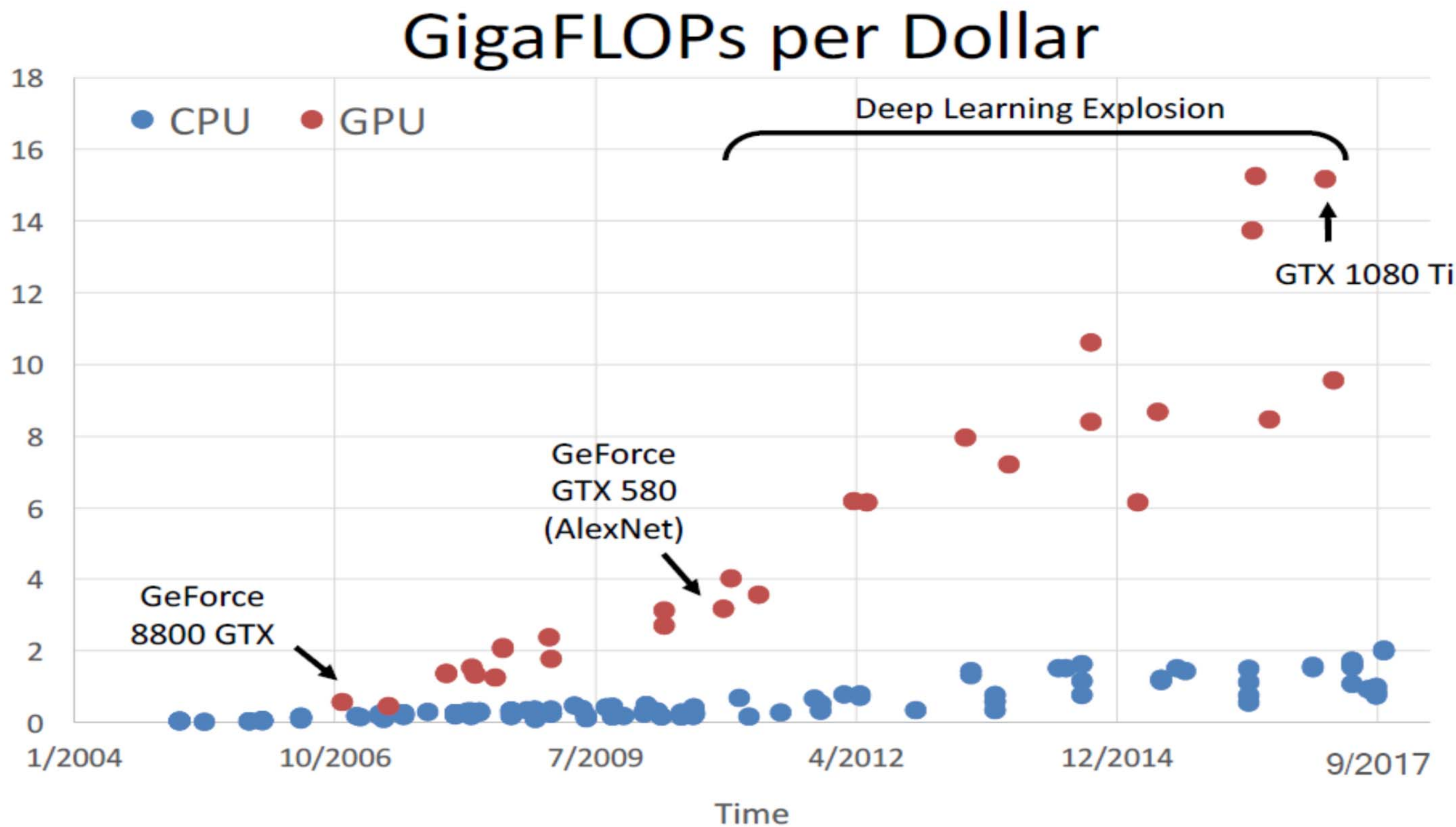
# of transistors GPUs

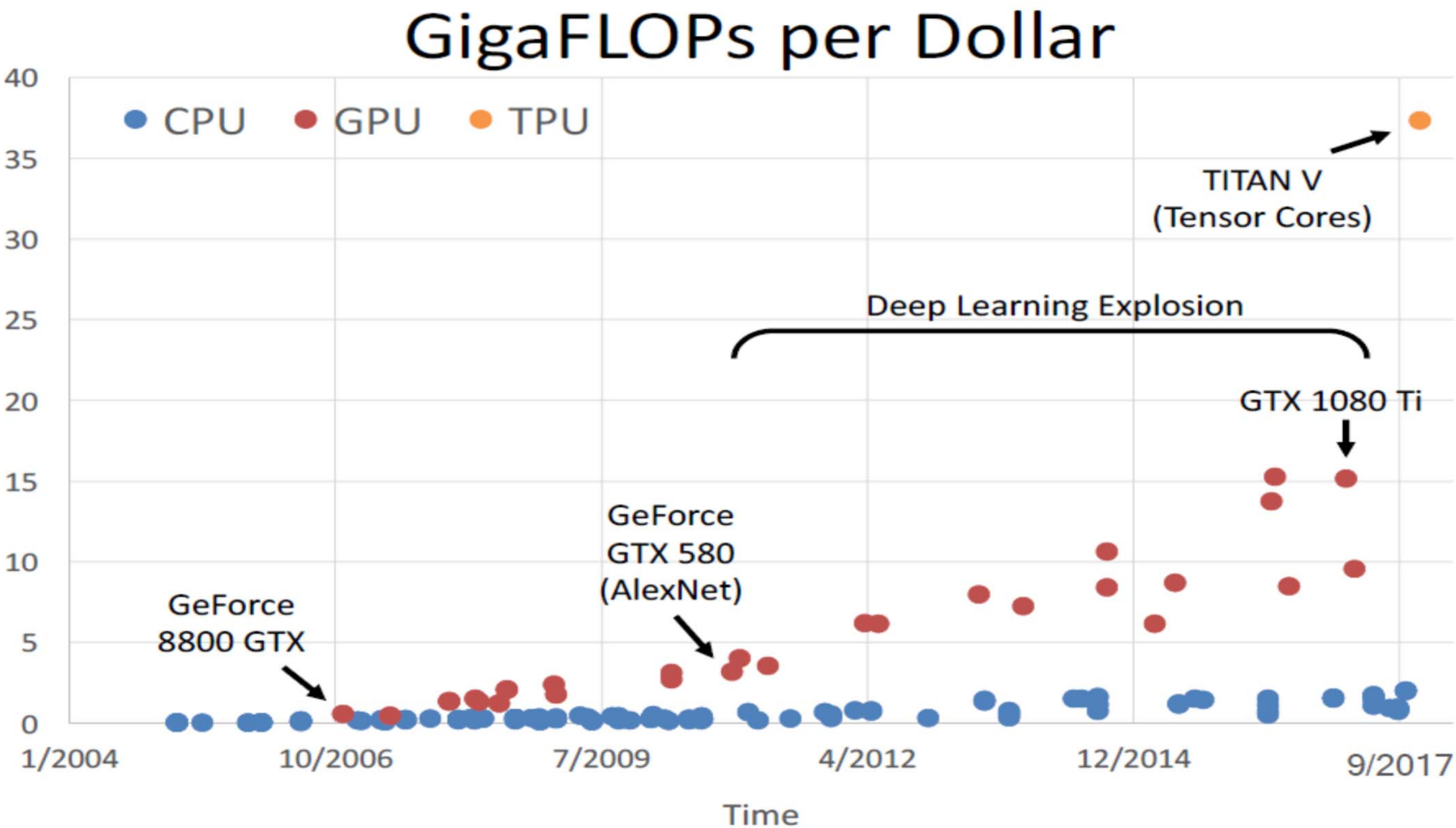
10^9

# of pixels used in training

10^{14} IMAGENET

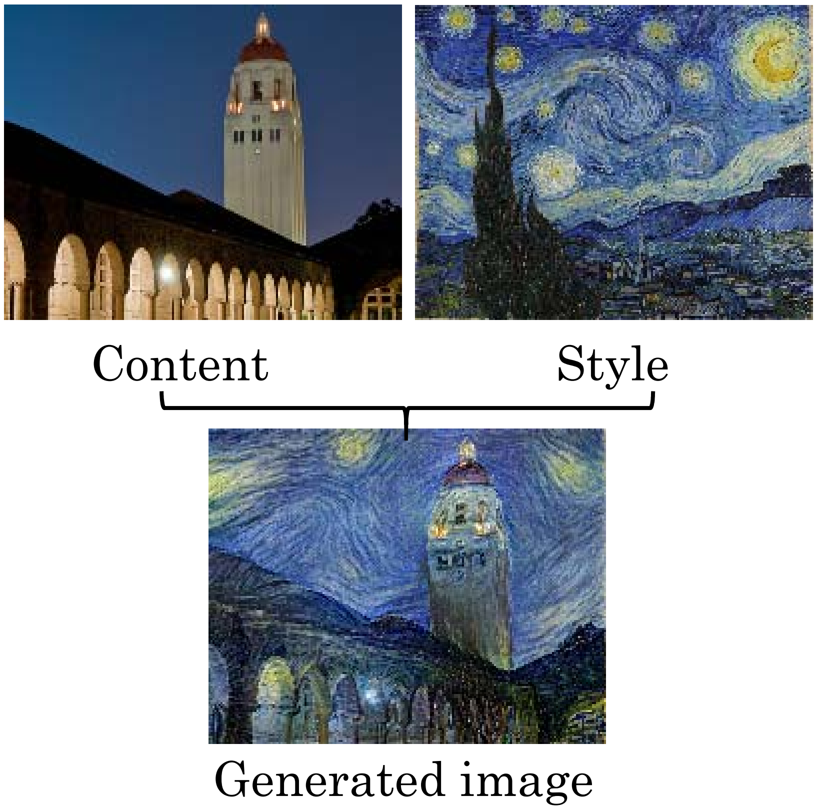
Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

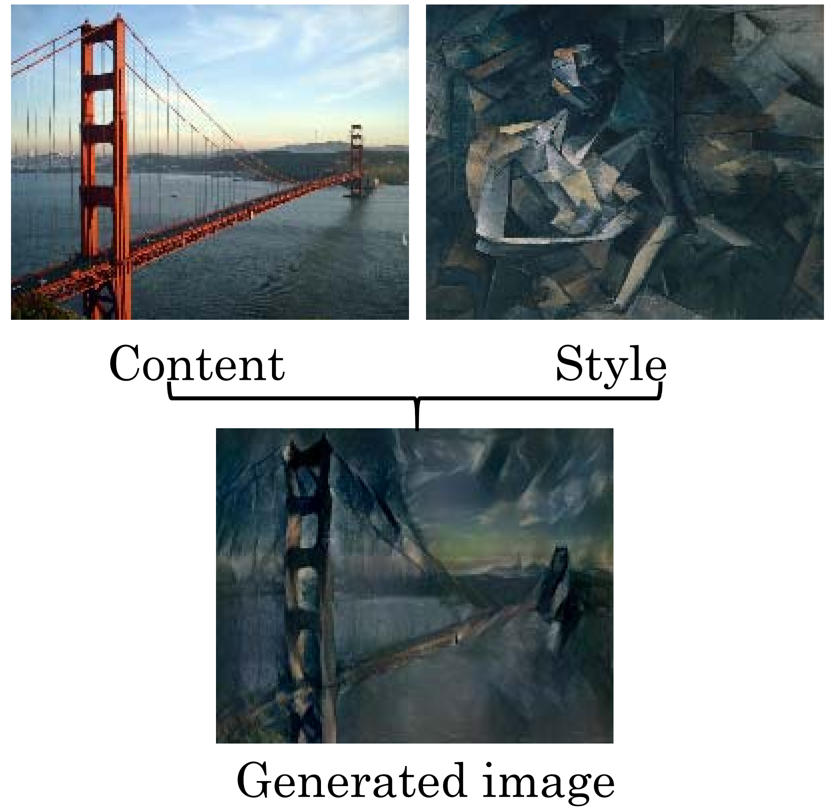

GigaFLOPs per Dollar
CPU
GPU
Deep Learning Explosion
GTX 1080 Ti
GeForce
GTX 580
(AlexNet)
GeForce
8800 GTX
18
16
14
12
10
8
6
4
2
0
1/2004
10/2006
7/2009
4/2012
12/2014
9/2017
Time



GigaFLOPs per Dollar
CPU
GPU
TPU
40
35
30
25
20
15
10
5
0
1/2004
10/2006
7/2009
4/2012
12/2014
9/2017
Time
TITAN V
(Tensor Cores)
Deep Learning Explosion
GTX 1080 Ti
GeForce
GTX 580
(AlexNet)
GeForce
8800 GTX

Neural style transfer



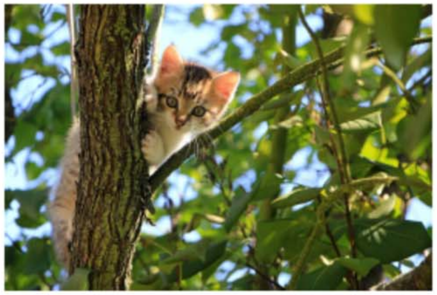


[Images generated by Justin Johnson]

Image Captioning: Example Results

Captions generated using neuraltalk2
All images are CC0 Public domain: cat suitcase, cat tree, dog, bear, surfers, tennis, giraffe, motorcycle

A cat sitting on a suitcase on the floor



A cat is sitting on a tree branch

A dog is running in the grass with a frisbee

A white teddy bear sitting in the grass

Two people walking on the beach with surfboards

A tennis player in action on the court

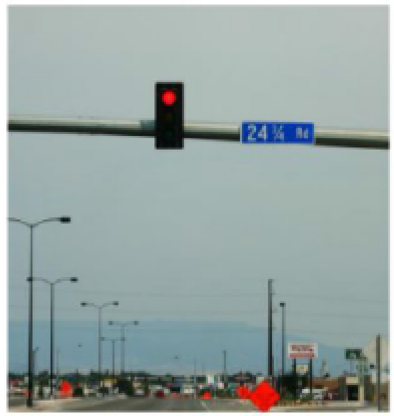
Two giraffes standing in a grassy field

A man riding a dirt bike on a dirt track

Visual Question Answering

Q: What endangered animal is featured on the truck?

A: **A bald eagle.**
A: A sparrow.
A: A humming bird.
A: A raven.



Q: Where will the driver go if turning right?

A: **Onto 24 ¾ Rd.**
A: Onto 25 ¾ Rd.
A: Onto 23 ¾ Rd.
A: Onto Main Street.

Q: Who is under the umbrella?

A: **Two women.**
A: A child.
A: An old man.
A: A husband and a wife.

Agrawal et al, "VQA: Visual Question Answering", ICCV 2015
Zhu et al, "Visual 7W: Grounded Question Answering in Images", CVPR 2016
Figure from Zhu et al, copyright IEEE 2016. Reproduced for educational purposes.

This image is licensed under CC BY-SA 3.0; changes made

Natural Language Processing

- Word Representation
- Word Embedding
- Sentiment Classification

Word representation

V = [a, aaron, …, zulu, <UNK>]

1-hot representation

$$\begin{array}{cccccc}
\text{Man} & \text{Woman} & \text{King} & \text{Queen} & \text{Apple} & \text{Orange} \\
(5391) & (9853) & (4914) & (7157) & (456) & (6257) \\
\begin{bmatrix}0\\0\\0\\0\\\vdots\\1\\\vdots\\0\\0\end{bmatrix} &
\begin{bmatrix}0\\0\\0\\0\\0\\\vdots\\1\\\vdots\\0\end{bmatrix} &
\begin{bmatrix}0\\0\\0\\\vdots\\1\\\vdots\\0\\0\\0\end{bmatrix} &
\begin{bmatrix}0\\0\\0\\0\\0\\\vdots\\1\\\vdots\\0\end{bmatrix} &
\begin{bmatrix}0\\\vdots\\1\\\vdots\\0\\0\\0\\0\\0\end{bmatrix} &
\begin{bmatrix}0\\0\\0\\0\\0\\\vdots\\1\\\vdots\\0\end{bmatrix}
\end{array}$$

I want a glass of orange ______.

I want a glass of apple______.

Featurized representation: word embedding

	Man (5391)	Woman (9853)	King (4914)	Queen (7157)	Apple (456)	Orange (6257)
			-0.95	0.97	0.00	0.01
			0.93	0.95	-0.01	0.00
			0.7	0.69	0.03	-0.02
			0.02	0.01	0.95	0.97

I want a glass of orange ______.

I want a glass of apple______.

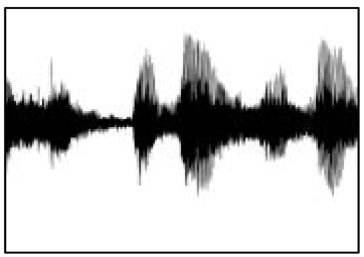
Sentiment classification problem

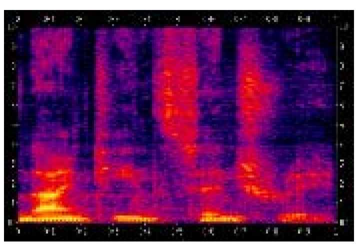
x	y
The dessert is excellent.	★★★★☆
Service was quite slow.	★★☆☆☆
Good for a quick meal, but nothing special.	★★★☆☆
Completely lacking in good taste, good service, and good ambience.	★☆☆☆☆

Speech recognition problem

x

audio clip





y

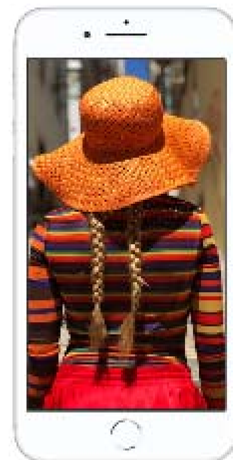
transcript

"the quick brown fox"

What is trigger word detection?

Amazon Echo
(Alexa)

Baidu DuerOS
(xiaodunihao)



Apple Siri
(Hey Siri)

Google Home
(Okay Google)

Generative Models

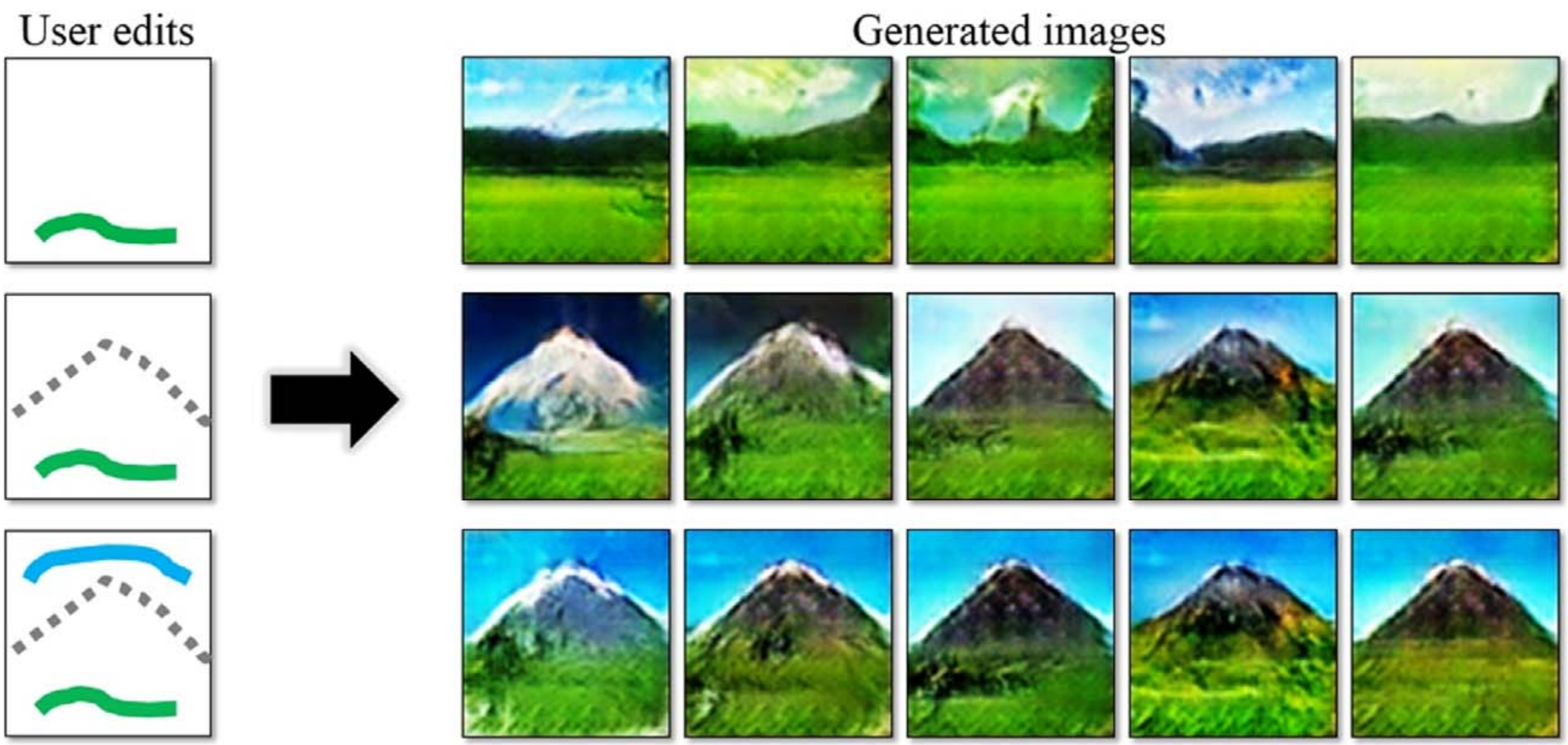
- Generative Adversarial Network
- Variational Autoencoder

Magic of GANs…

Which one is Computer generated?

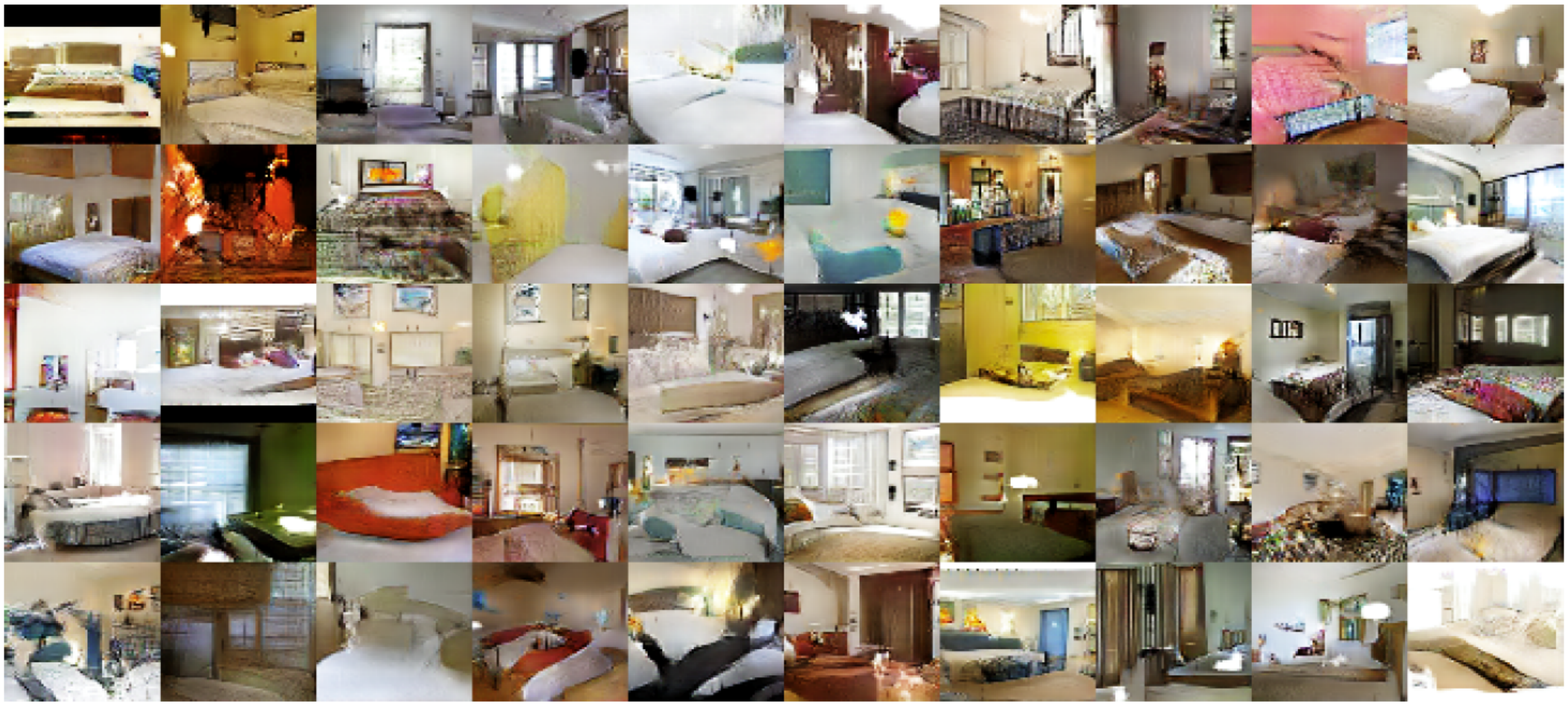
Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network." *arXiv preprint arXiv:1609.04802* (2016).

Magic of GANs…



http://people.eecs.berkeley.edu/~junyanz/projects/gvm/

DCGAN: Bedroom images



Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv:1511.06434 (2015).

Image-to-Image Translation

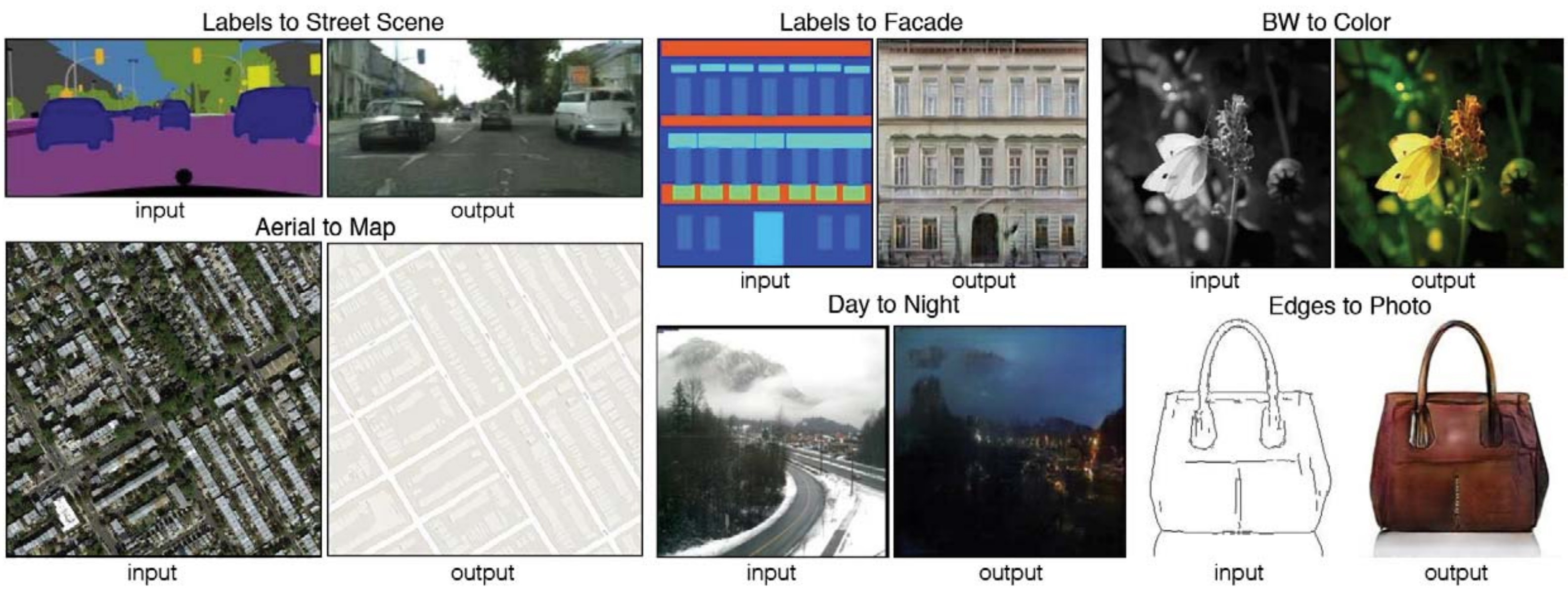


Figure 1 in the original paper.

[Link to an interactive demo of this paper]

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. "**Image-to-image translation with conditional adversarial networks**". arXiv preprint arXiv:1611.07004. (2016).

Text-to-Image Synthesis

Motivation

Given a text description, generate images closely associated.

Uses a conditional GAN with the generator and discriminator being condition on “dense” text embedding.

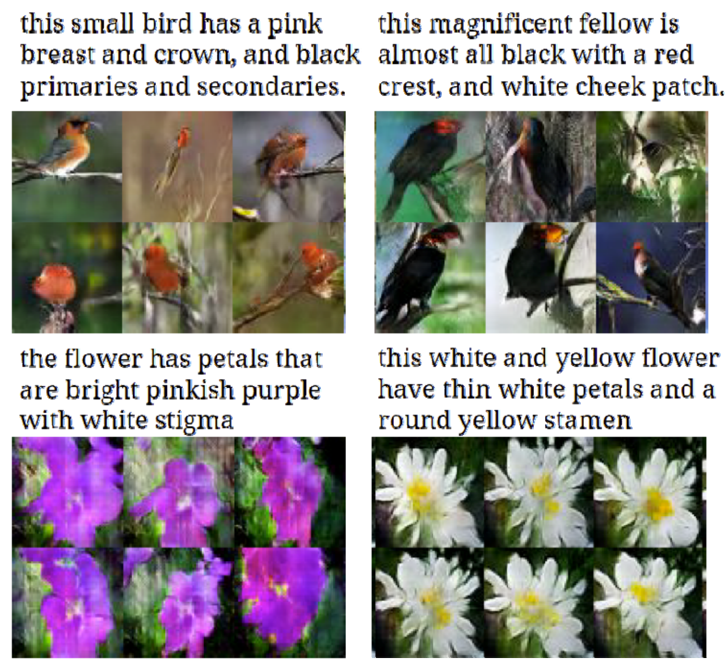
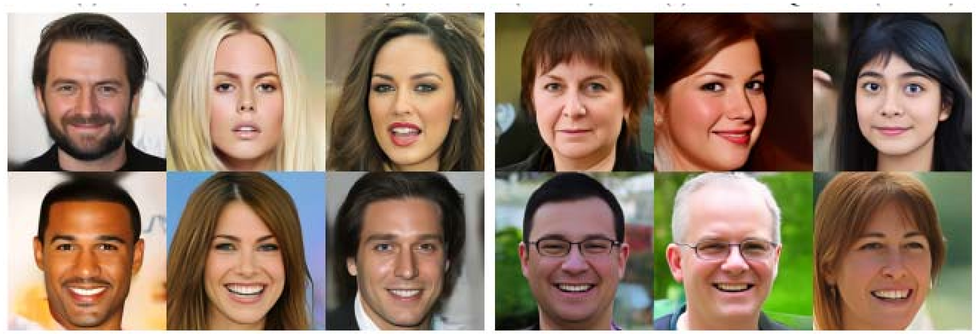


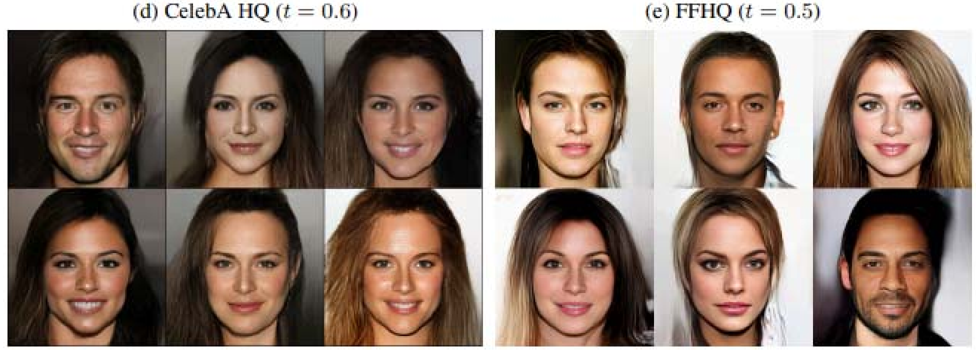
Figure 1 in the original paper.

Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., & Lee, H. “**Generative adversarial text to image synthesis**”. ICML (2016).

Face images generated with a Variational Autoencoder



(d) CelebA HQ ($t = 0.6$) (e) FFHQ ($t = 0.5$)



A. Vahdat and J. Kautz, “NVAE: A deep hierarchical variational autoencoder,” in Proc. Conf. Neural Inf. Process. Syst., 2020, pp. 19667–19679

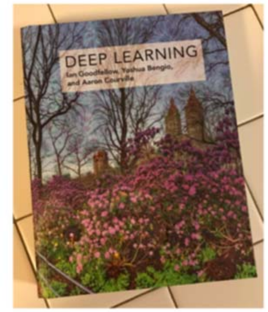
Related Courses

- Deep Learning, Andrew NG.

- Convolutional Neural Networks for Visual Recognition, Fei-Fei Lee, Justin Johnson, Serena Yeung

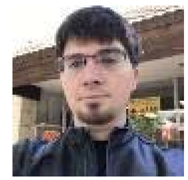
- Neural Network for Machine Learning

Textbooks



- Deep learning, I. Goodfellow, j. Bengio

MIT Press, 2016



- Neural Networks and deep learning, M. Nielsen

Determination Press, 2015

Teacher Assistants

- Fatemeh Piri

fatemeh_piri@ec.iut.ac.ir

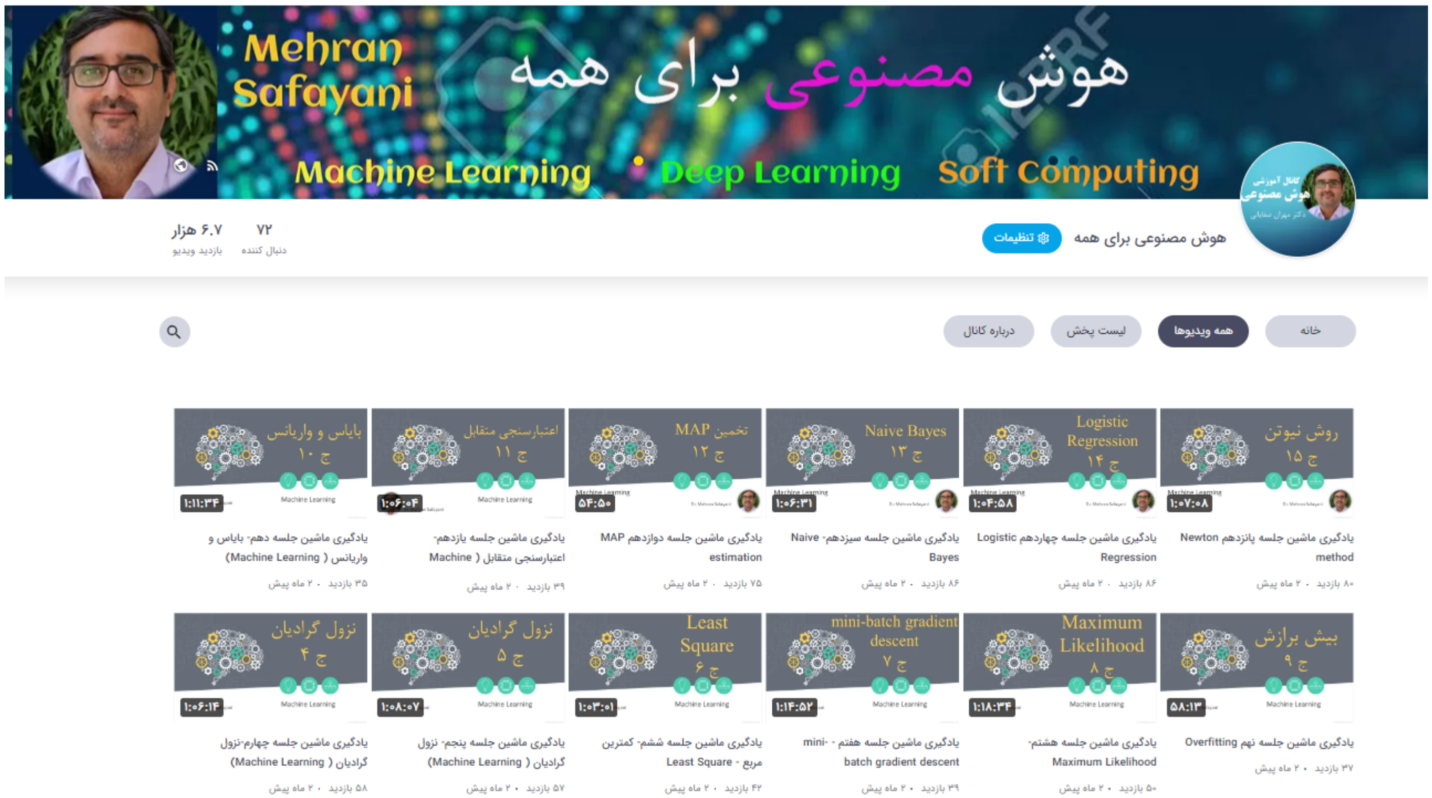
- Masomeh sadat Razavi

masoomehsadat.razavi@gmail.com

Course Strategy

- Assignments and project 20%-30% (python)
- Midterm 30%-40% marks
- Final 30%-40% marks

برای دریافت آموزش های بیشتر در حوزه یادگیری عمیق و یادگیری ماشین کانال زیر را دنبال کنید:

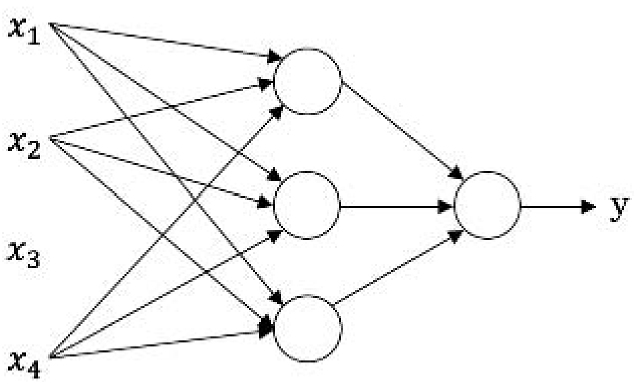


https://www.aparat.com/mehran.safayani

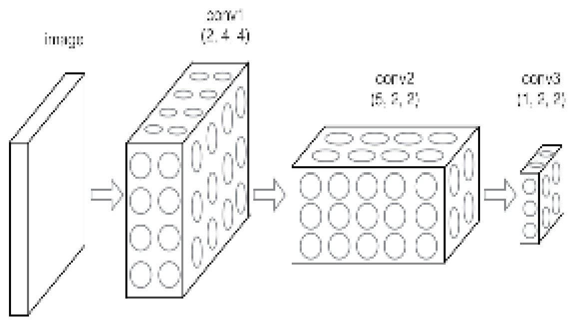
Supervised Learning

Input(x)	Output (y)	Application
Home features	Price	Real Estate
Ad, user info	Click on ad? (0/1)	Online Advertising
Image	Object (1,…,1000)	Photo tagging
Audio	Text transcript	Speech recognition
English	Chinese	Machine translation
Image, Radar info	Position of other cars	Autonomous driving

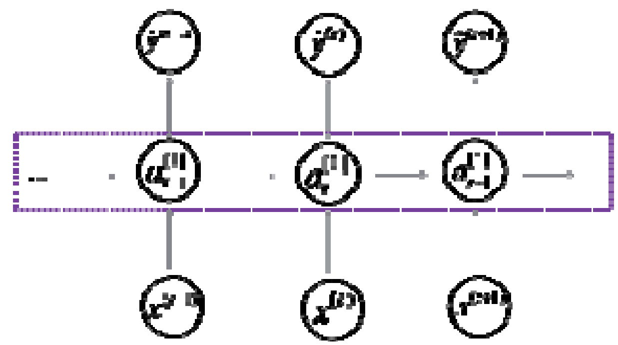
Neural Network examples



Standard NN



Convolutional NN



Recurrent NN

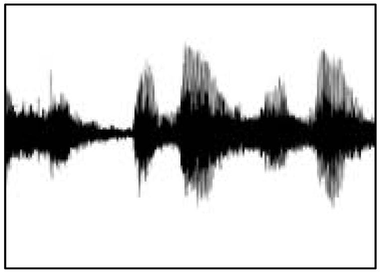
Supervised Learning

Structured Data

Size	#bedrooms	···	Price (1000$s)
2104	3		400
1600	3		330
2400	3		369
⋮	⋮		⋮
3000	4		540

User Age	Ad Id	···	Click
41	93242		1
80	93287		0
18	87312		1
⋮	⋮		⋮
27	71244		1

Unstructured Data



Audio

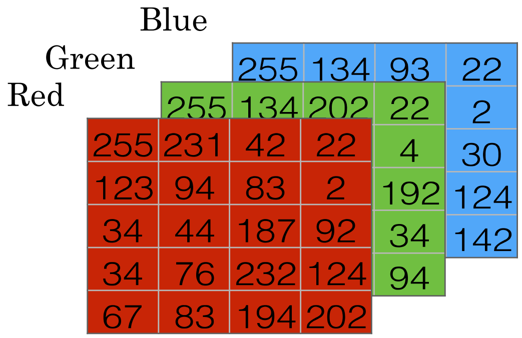
Image

Four scores and seven years ago…

Text

Binary Classification

1 (cat) vs 0 (non cat)



Gradient Descent

Recap: $\hat{y} = \sigma(w^T x + b)$, $\sigma(z) = \frac{1}{1+e^{-z}}$

$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m}\mathcal{L}(\hat{y}^{(i)}, y^{(i)}) = \frac{1}{m}\sum_{i=1}^{m} y^{(i)}\log\hat{y}^{(i)} + (1-y^{(i)})\log(1-\hat{y}^{(i)})$$

Want to find w, b that minimize $J(w,b)$