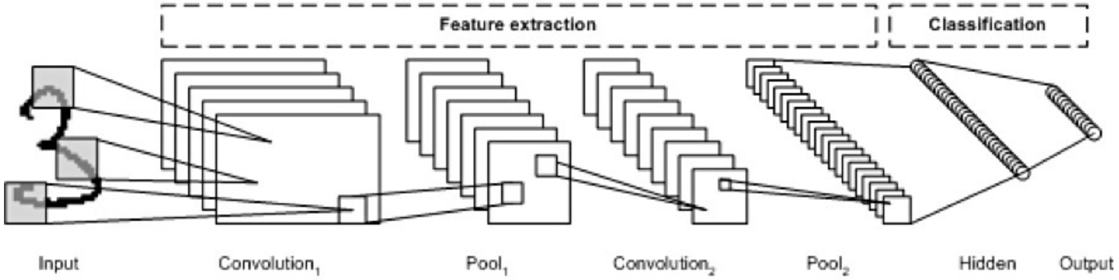
Computational vision:

Convolutional Neural Networks

Class 10: Visión Artificial



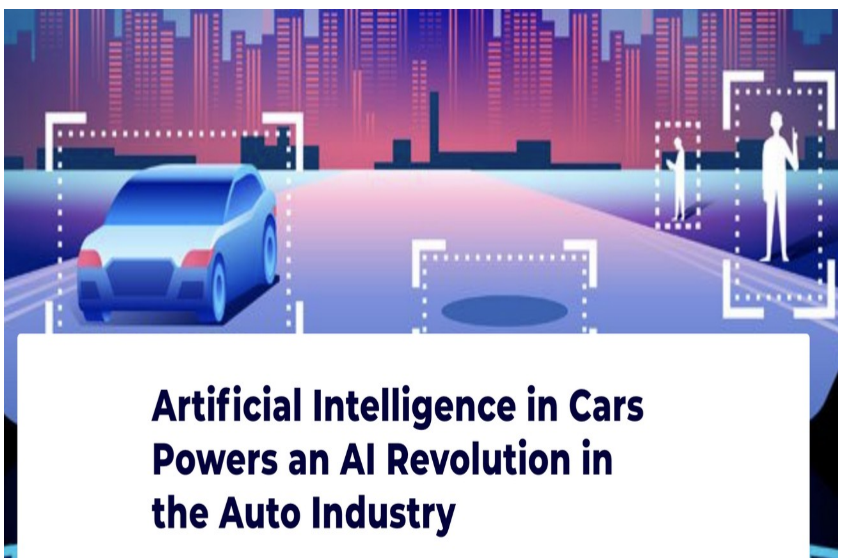
Petia Radeva

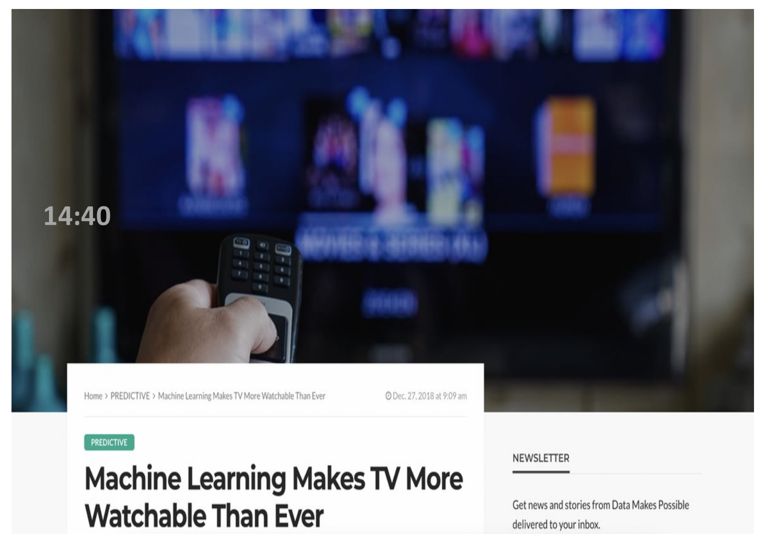
2023-2024

Index

- AI, Machine learning & Deep learning
- What is a Convolutional Neural Network?
  - Layers
  - Optimization
- Applications

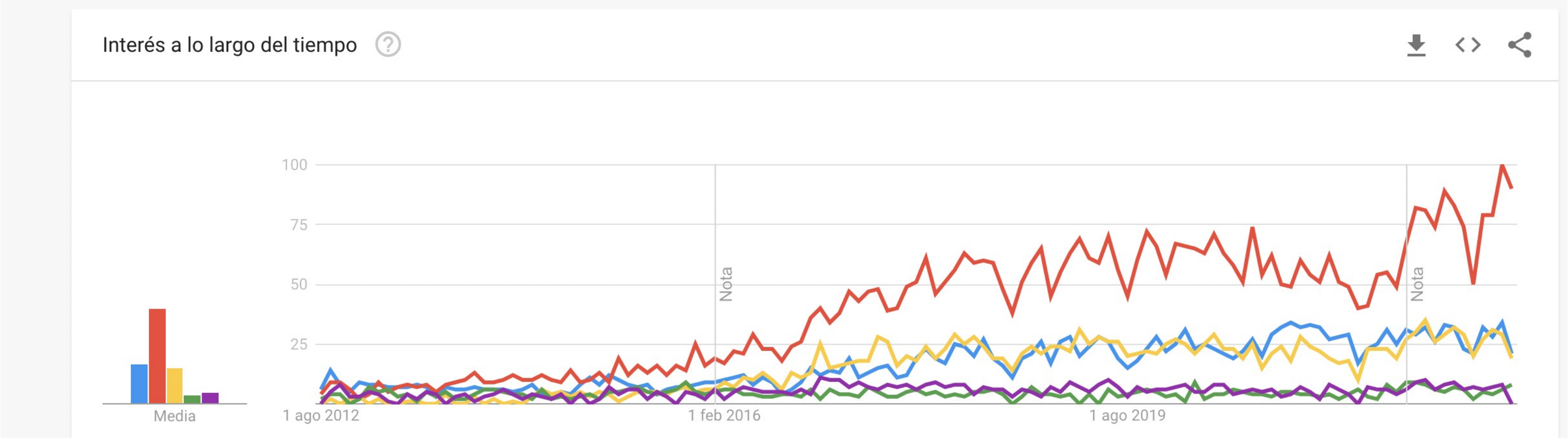
Deep learning everywhere





Artificial Intelligence

- Machine Learning
- Artificial Intelligence
- Deep learning
- Neural networks
- Computer Vision



Google Scholar reveals its most influential papers

1. **"Deep Residual Learning for Image Recognition" (2016)** *Proceedings of the IEEE/C Computer Vision and Pattern Recognition* 25,256 citations

2. **"Deep learning" (2015)** *Nature* 16,750 citations

3. **"Going Deeper with Convolutions" (2015)** *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 14,424 citations

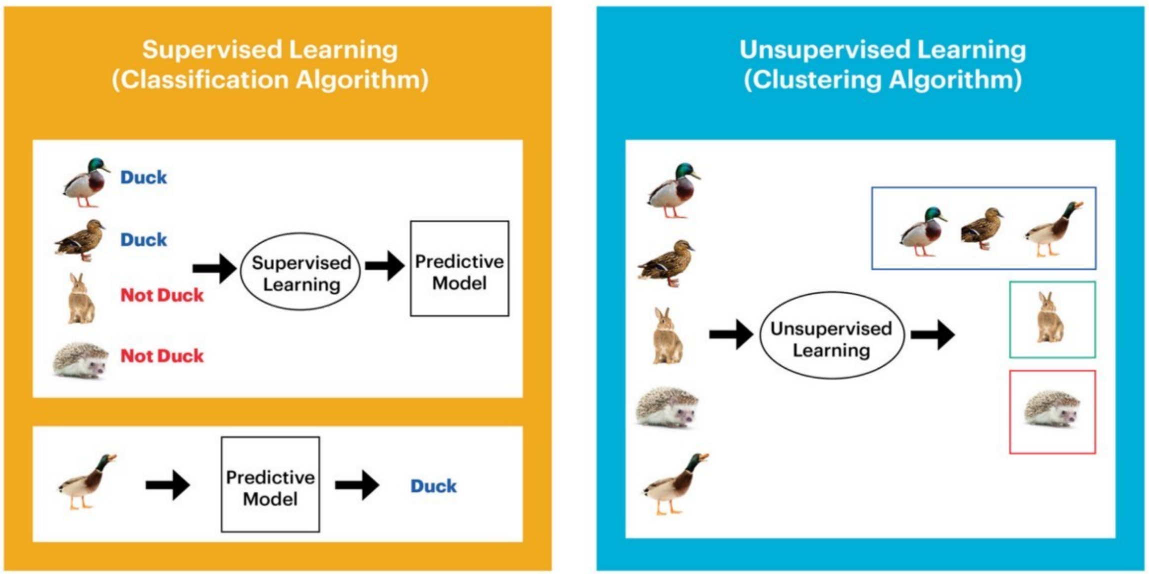
4. **"Fully Convolutional Networks for Semantic Segmentation" (2015)** *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition* 10,153 citations

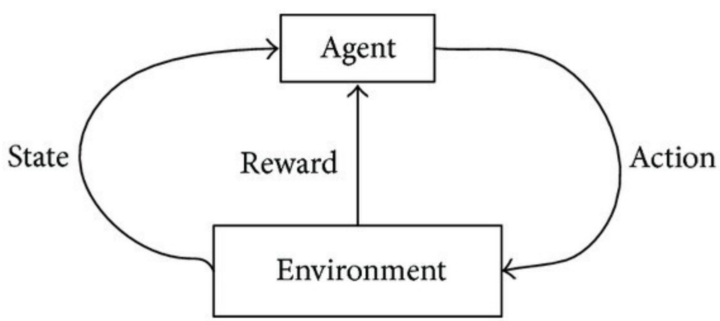
5. **"Prevalence of Childhood and Adult Obesity in the United States, 2011-2012" (2014)** *JAMA* 8,057 citations

6. **"Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013" (2014)** *Lancet* 7,371 citations

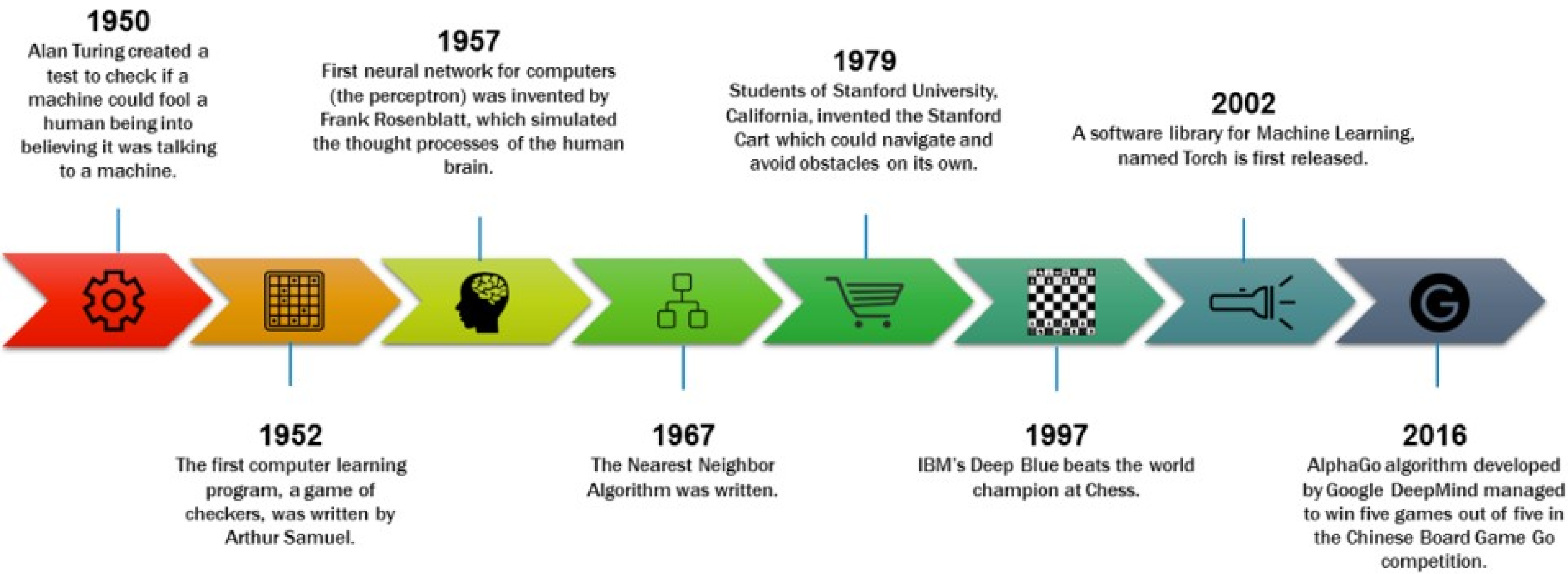
7. **"Observation of Gravitational Waves from a Binary Black Hole Merger" (2016)** *Physical Review Letters* 6,009 citations

Supervised vs. Unsupervised vs Reinforcement Learning





Ai, ML and DL evolution



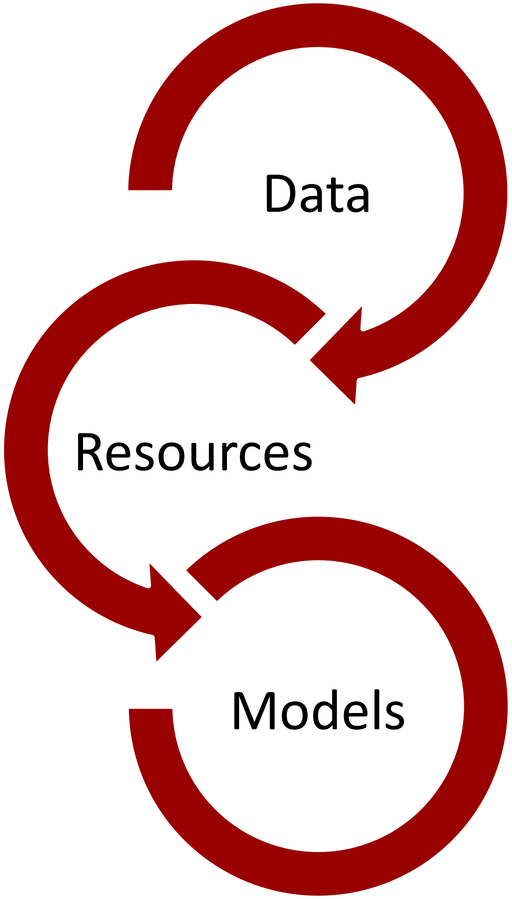
Deep learning – where does it come from?

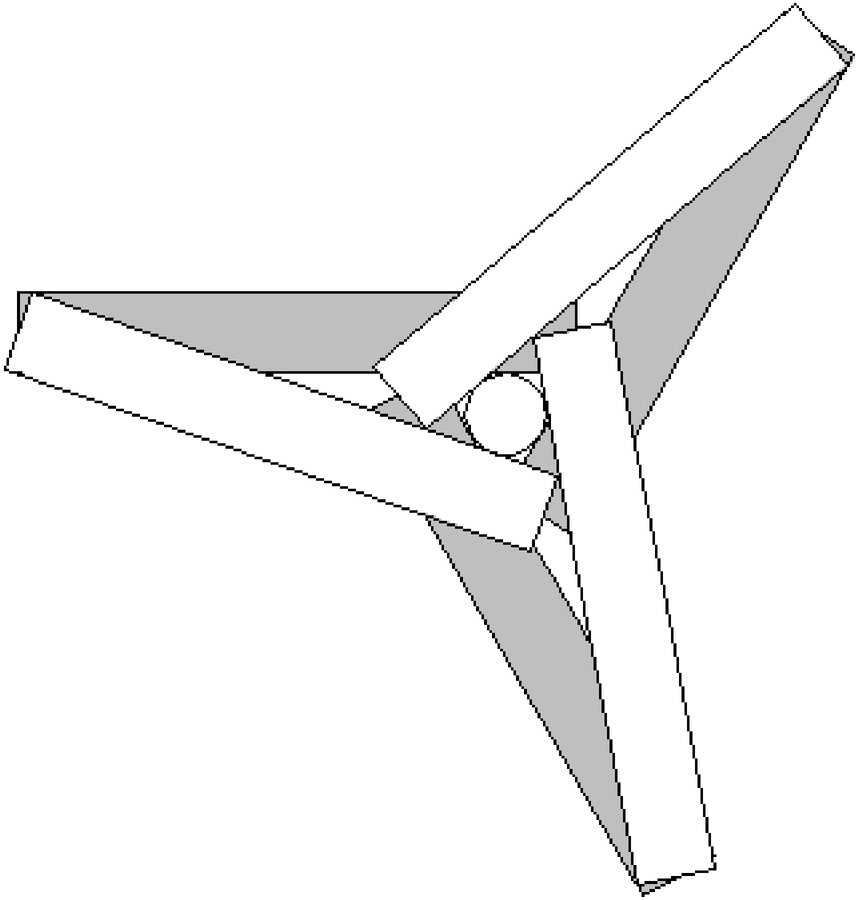
- **1943:** neurophysiologist Warren McCulloch and mathematician Walter Pitts – a **neuronal model** using an electrical circuit
- **1950**: Alan Turing created the world-famous **Turing Test**: a computer to pass, has to be able to convince a human that it is a human and not a computer.
- **1952**: Arthur Samuel created the first computer **program which could learn** as it ran. It was a game which played checkers.
- **1958**: Frank Rosenblatt designed the first artificial neural network, called **Perceptron**. The main goal of this was pattern and shape recognition.
- **1959**, Bernard Widrow and Marcian Hoff created
  - ADELINE to detect binary patterns (in a stream of bits, to predict the next one).
  - MADELINE to eliminate echo on phone lines, still in use today.
- **1986**: Neural networks use **back propagation** allowing multiple layers to be used in a neural network, creating what are known as ‘slow learners’.
- Late **1980s** and **1990s** did not bring much to the field.
- **1997**, the IBM computer Deep Blue (a chess-playing computer) beat the world chess champion.
- **1998** AT&T Bell Laboratories on digit recognition resulted in good accuracy in detecting handwritten postcodes from the US Postal Service.

Early History of Machine Learning

What changed today?

The magic triangle



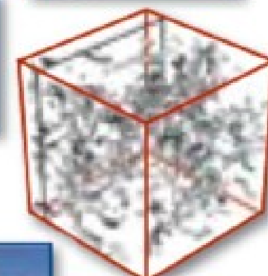


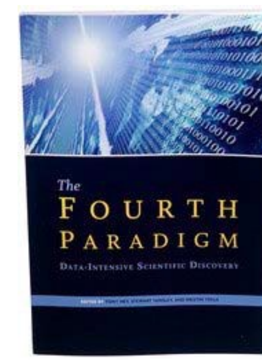
The Jim Cray's paradigms

Science Paradigms

- Thousand years ago:
  science was **empirical**
  describing natural phenomena
- Last few hundred years:
  theoretical branch
  using models, generalizations
- Last few decades:
  a **computational** branch
  simulating complex phenomena
- Today: **data exploration** (eScience)
  unify theory, experiment, and simulation
  - Data captured by instruments or generated by simulator
  - Processed by software
  - Information/knowledge stored in computer
  - Scientist analyzes database/files using data management and statistics

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi Gp}{3} - K\frac{c^2}{a^2}$$





Toni Hey, 2009

Data

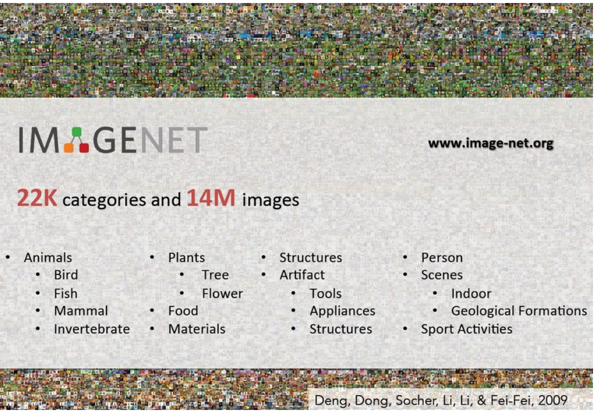
90% of all digital data were generated last 2 years.

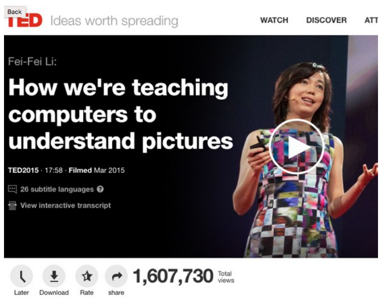
Every minute of the day:

- 4M YouTube videos watched
- 456K tweets on Twitter
- 46K potos posted in Instagram
- 16M text messages sent
- <u>103M spam emails</u> sent

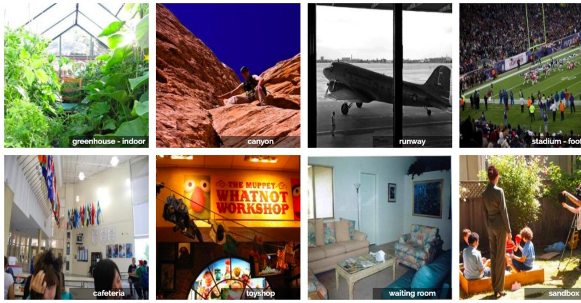
https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#46be238160ba

DL Datasets

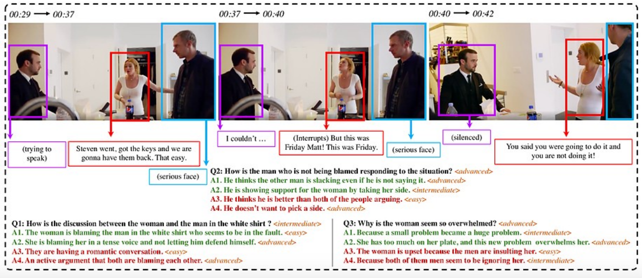




LVIS Challenge:
2.2M masks, 16K images

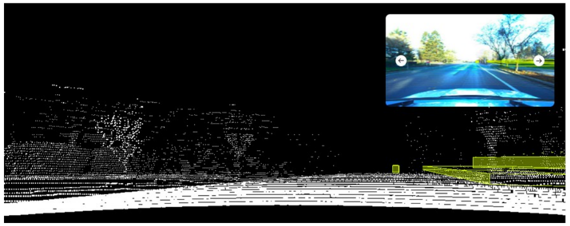


Places2: 10M images



SocialIQ

TACO: Waste in the wild



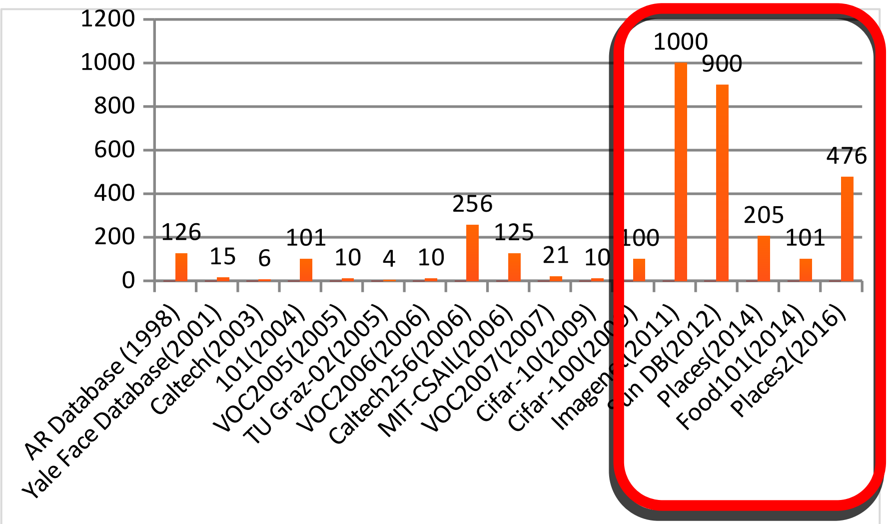
[Lyft Level 5]

FastMRI

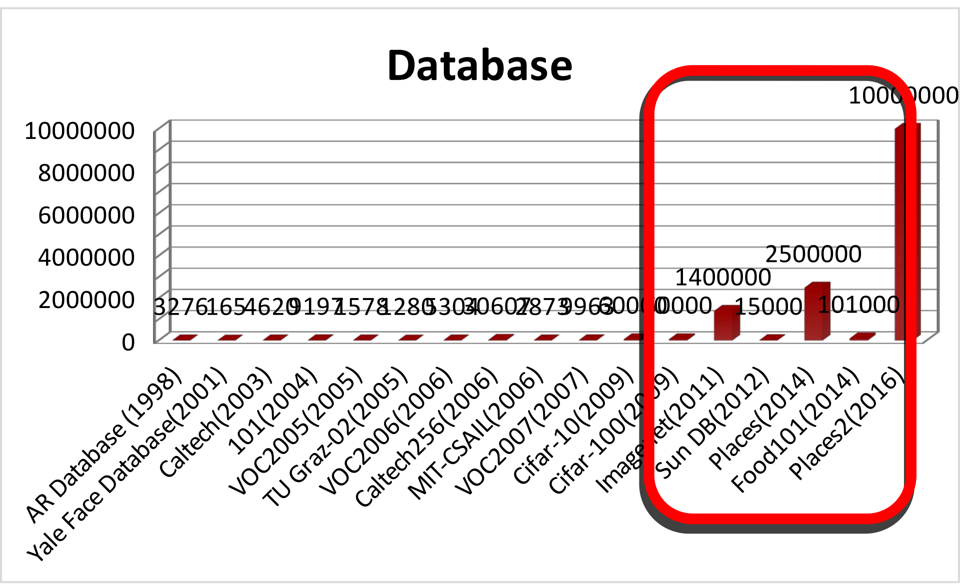
https://www.datasetlist.com

Image databases evolution

Number of objects/Database

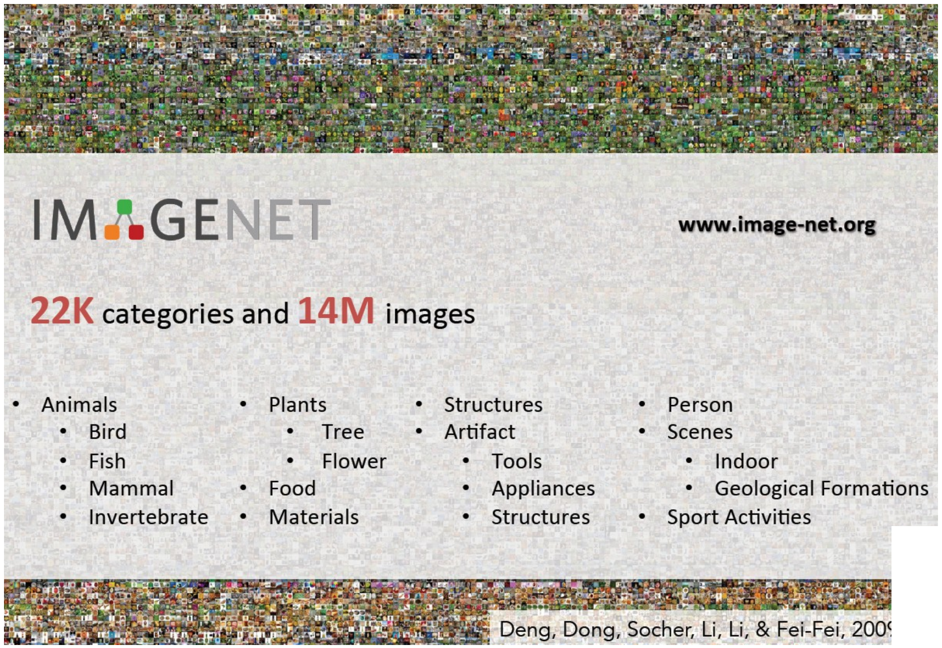


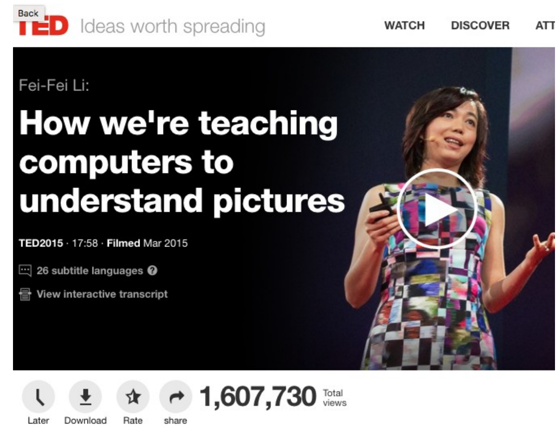
Number of images/Database

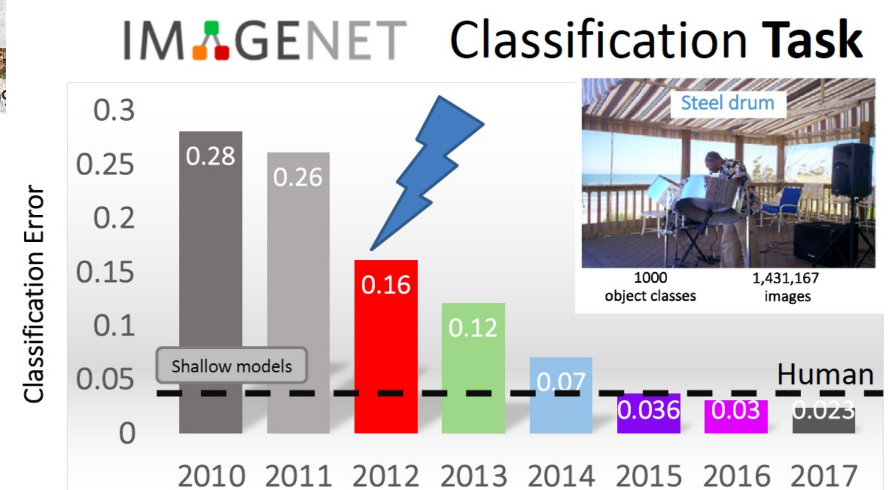


ImageNet & Deep learning

Imagenet



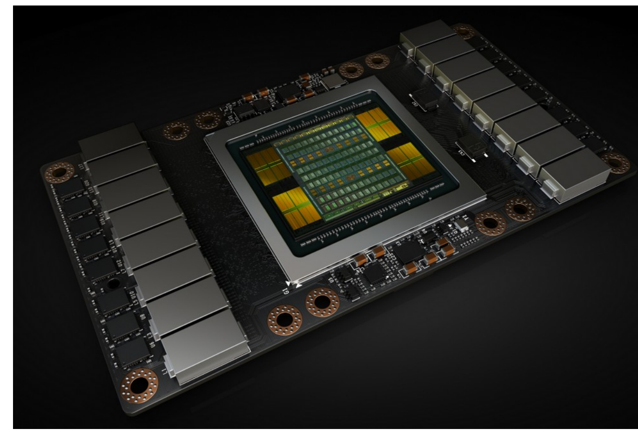




Deng et al. *CVPR*, 2009; Russakovsky et al. *IJCV*, 2012;

The Importance of GPUs

- Nvidia Tensor Cores - 2017
- Google Tensor Processing Unit (TPU) - 2016
- Intel - Nervana Neural Processor - 2017
- GPUs in Cloud Computing (Google, 2017)



$$\underbrace{\mathbf{D}}_{\text{FP16 or FP32}} = \underbrace{\begin{pmatrix} A_{0,0} & A_{0,1} & A_{0,2} & A_{0,3} \\ A_{1,0} & A_{1,1} & A_{1,2} & A_{1,3} \\ A_{2,0} & A_{2,1} & A_{2,2} & A_{2,3} \\ A_{3,0} & A_{3,1} & A_{3,2} & A_{3,3} \end{pmatrix}}_{\text{FP16}} \underbrace{\begin{pmatrix} B_{0,0} & B_{0,1} & B_{0,2} & B_{0,3} \\ B_{1,0} & B_{1,1} & B_{1,2} & B_{1,3} \\ B_{2,0} & B_{2,1} & B_{2,2} & B_{2,3} \\ B_{3,0} & B_{3,1} & B_{3,2} & B_{3,3} \end{pmatrix}}_{\text{FP16}} + \underbrace{\begin{pmatrix} C_{0,0} & C_{0,1} & C_{0,2} & C_{0,3} \\ C_{1,0} & C_{1,1} & C_{1,2} & C_{1,3} \\ C_{2,0} & C_{2,1} & C_{2,2} & C_{2,3} \\ C_{3,0} & C_{3,1} & C_{3,2} & C_{3,3} \end{pmatrix}}_{\text{FP16 or FP32}}$$

GPU cores is based on matrix multiplication

https://www.doc.ic.ac.uk/~jce317/history-machine-learning.html#top

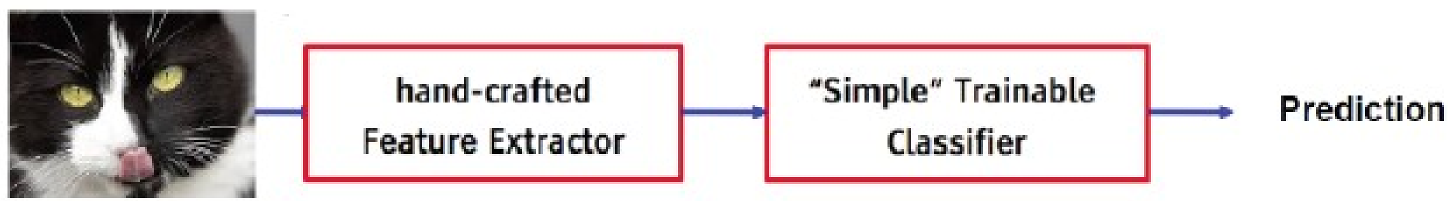
What is a Neural Network?

Index

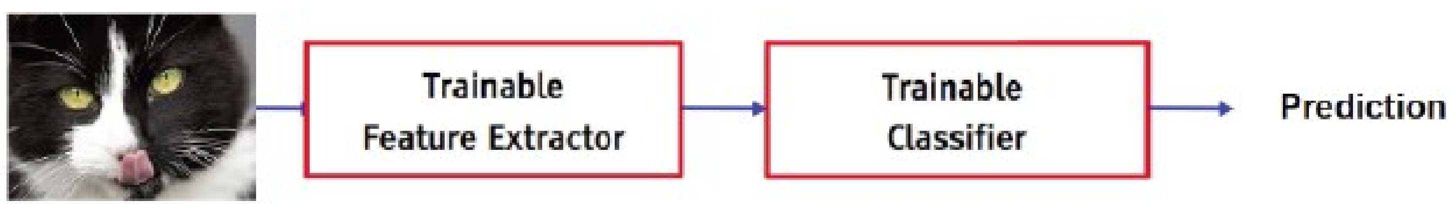
- AI, Machine learning & Deep learning
- What is a Convolutional Neural Network?
  - Layers
  - Optimization
- Applications

Why Deep learning?

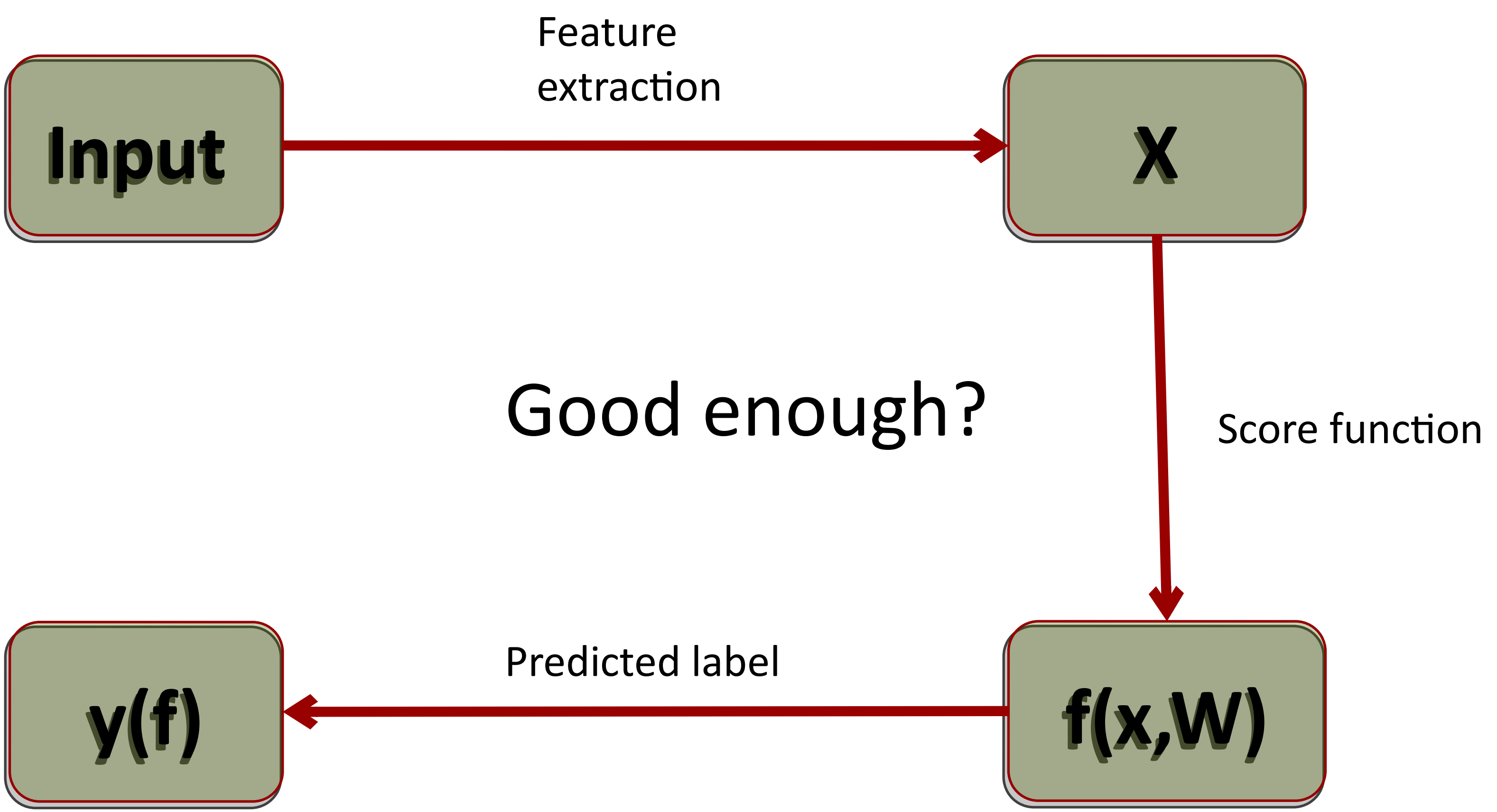
Classical way of solving Computer Vision problems



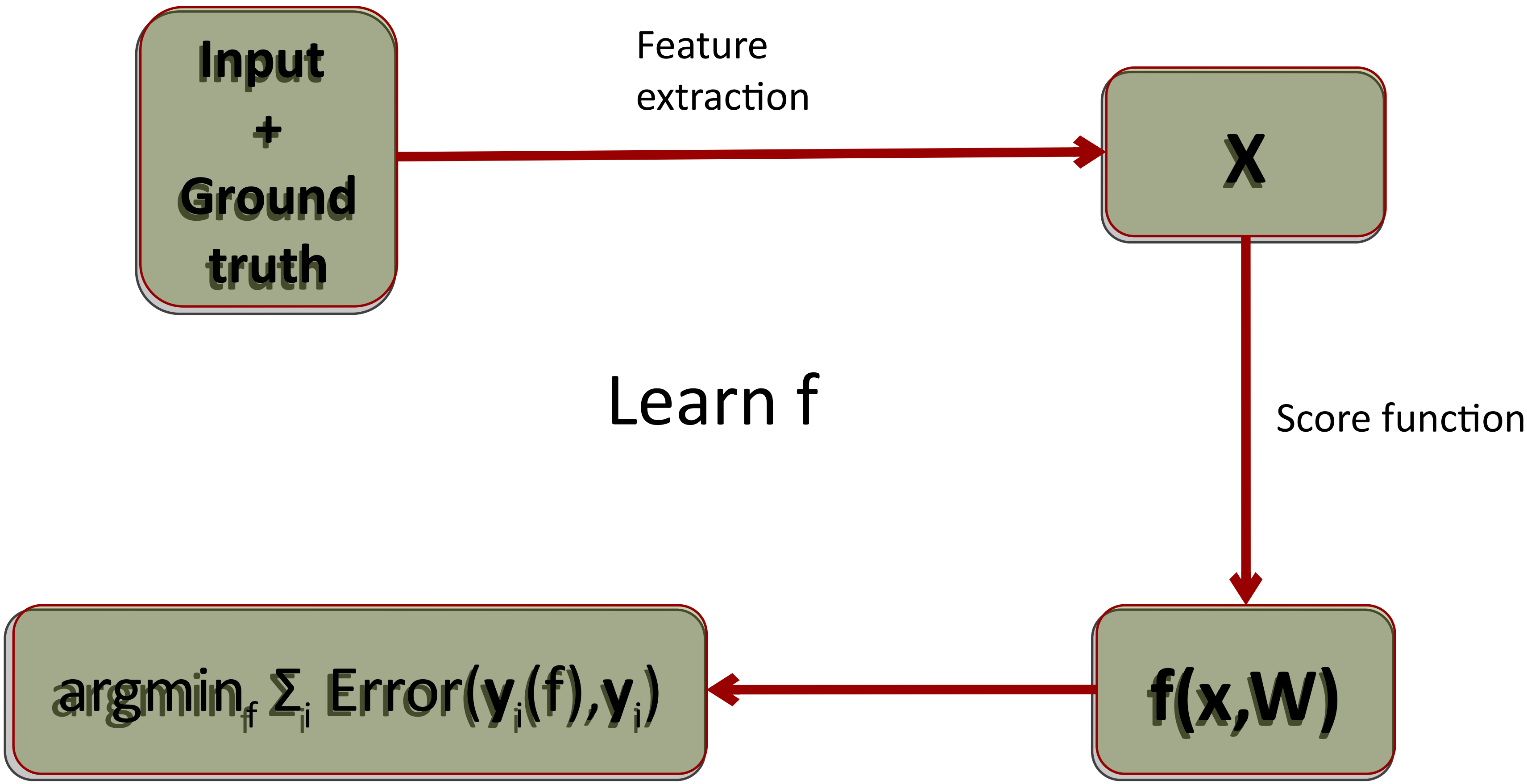
How Computer Vision problems are solved by Deep Learning



The learning pipeline



The traning process



The learning process

Training data $\{(\mathbf{x}_i,\mathbf{y}_i),\ i = 1,2,\ldots,n\}$

Measure of prediction quality (error, loss)

$$\operatorname{argmin}_f \Sigma_i \operatorname{Error}(\mathbf{y}_i(f),\mathbf{y}_i)$$

Expectation over data distribution

Prediction

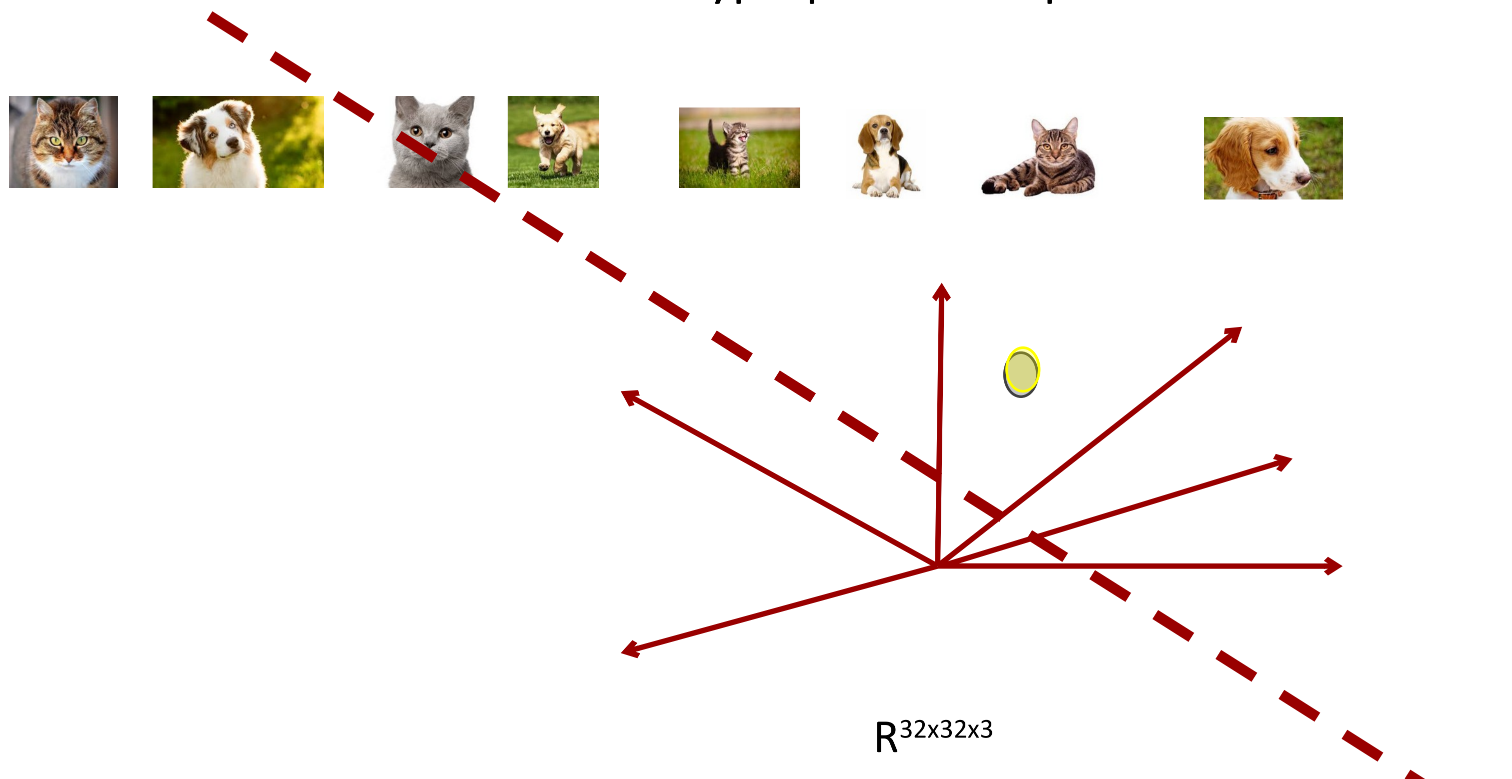
Ground Truth

Loss function the negative conditional log-likelihood, with the interpretation that $f_i(X)$ estimates $P(Y=i|X)$:

$$\mathbf{L(f(x),y)) = -log\ f_i(x),\ where\ f_i(x)>=0,\ \ \Sigma_i\ f_i(x) = 1.}$$

Linear classification

Given two classes how to learn a hyperplane to separate them?

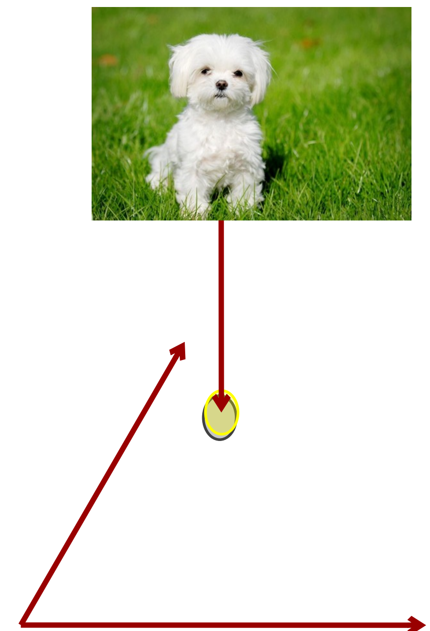


To find the hyperplane that separates dogs from cats, we need to define:

- The score function
- The loss function
- And the optimization process.

How to project data in the feature space:

$$\overset{3\text{x}1}{f(x)}=\underset{3\text{x}3072}{W}\ \overset{3072\text{x}1}{x} + \underset{3\text{x}1}{b}$$



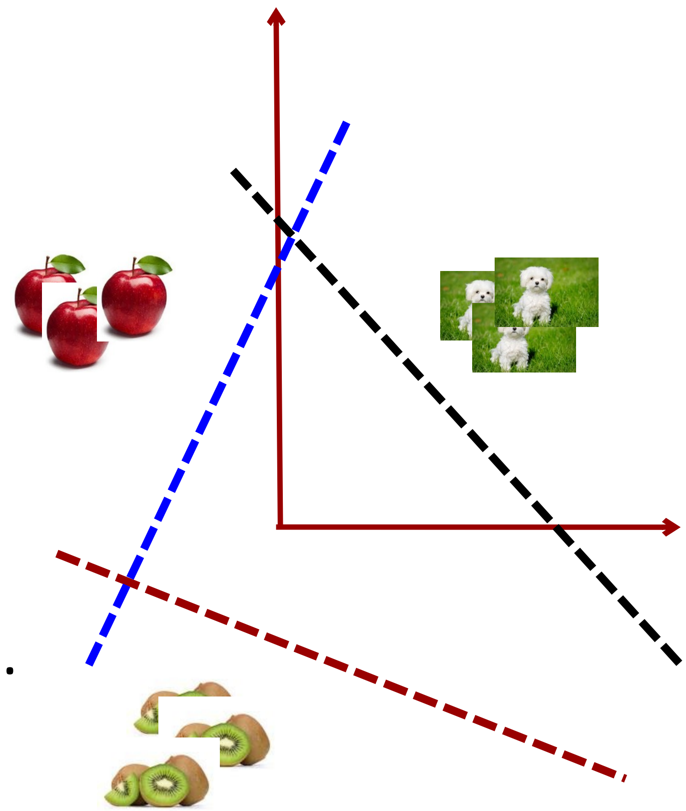
If x is an image of (32x32x3), -> x in $R^{3072,}$

The matrix W is (3x3072).

The bias vector b is 3-dimensional.

How to project data in the feature space:

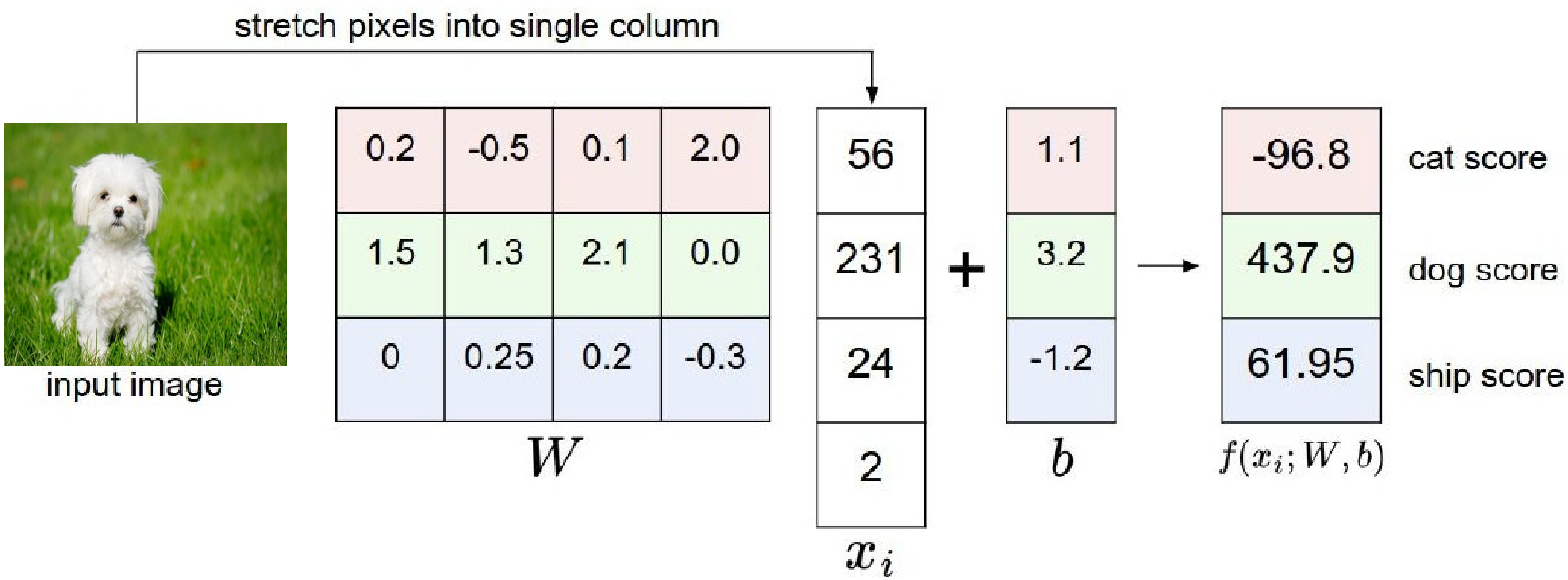
$$\underbrace{f(x)}_{3\text{x}1} = \underbrace{W}_{3\text{x}3072}\ \underbrace{x}_{3072\text{x}1} + \underbrace{b}_{3\text{x}1}$$



If we have 3 classes, f(x) will give 3 scores.

Image classification

Example with an image with 4 pixels, and 3 classes (cat/dog/ship)



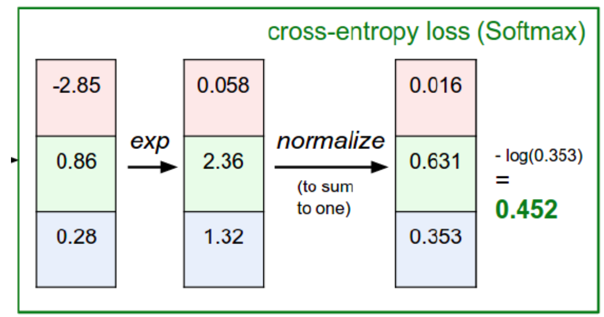
Adapted from: Fei-Fei Li & Andrej Karpathy & Justin Johnson

Loss function and optimisation

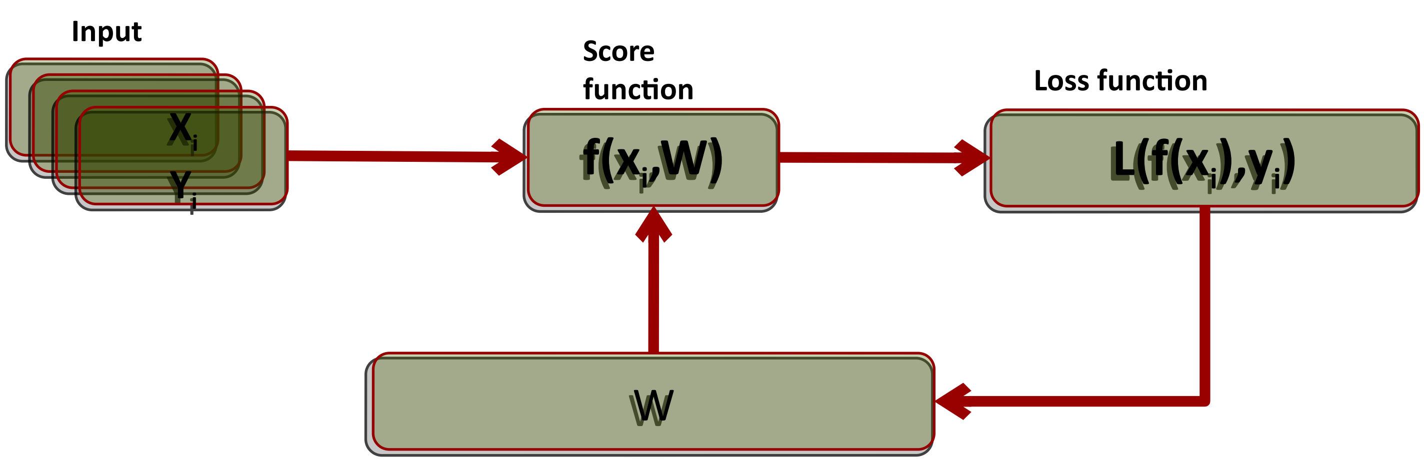
- **Question:** if you were to assign a single number to how unhappy you are with these scores, what would you do?

$$L_i = -\log\left(\boxed{\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}}\right)$$

softmax function

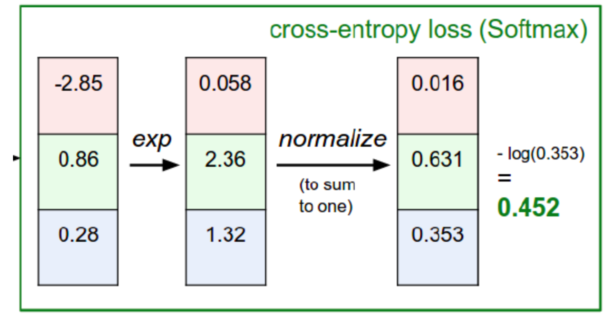


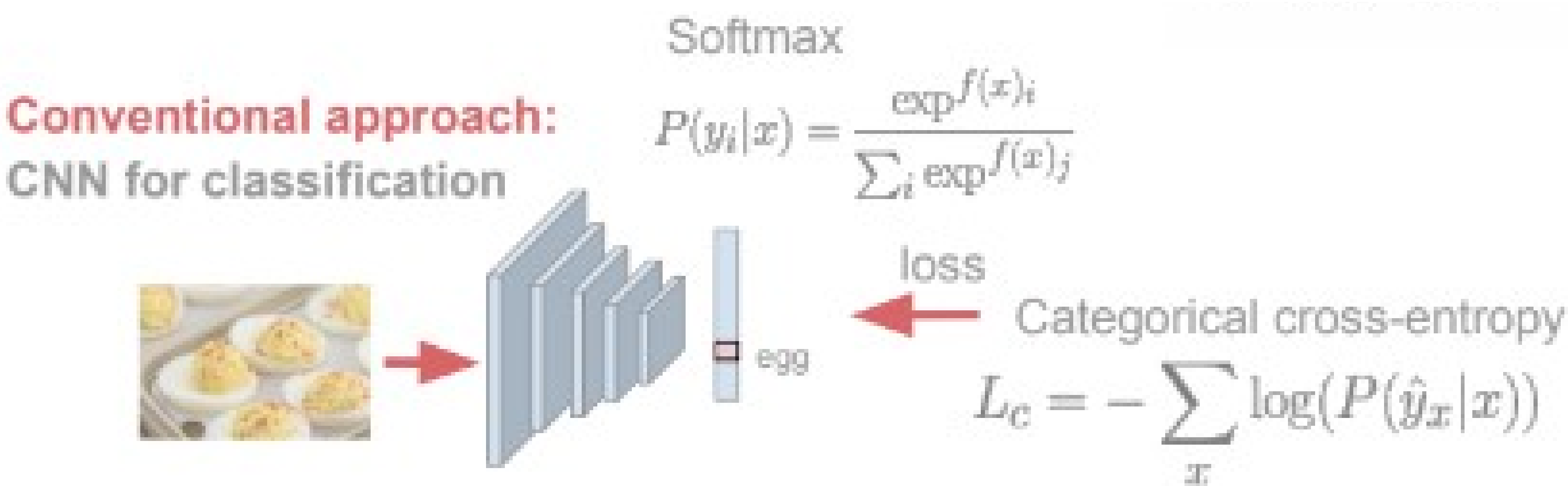
Question : Given the score and the loss function, how to find the parameters W?



Single-label classification

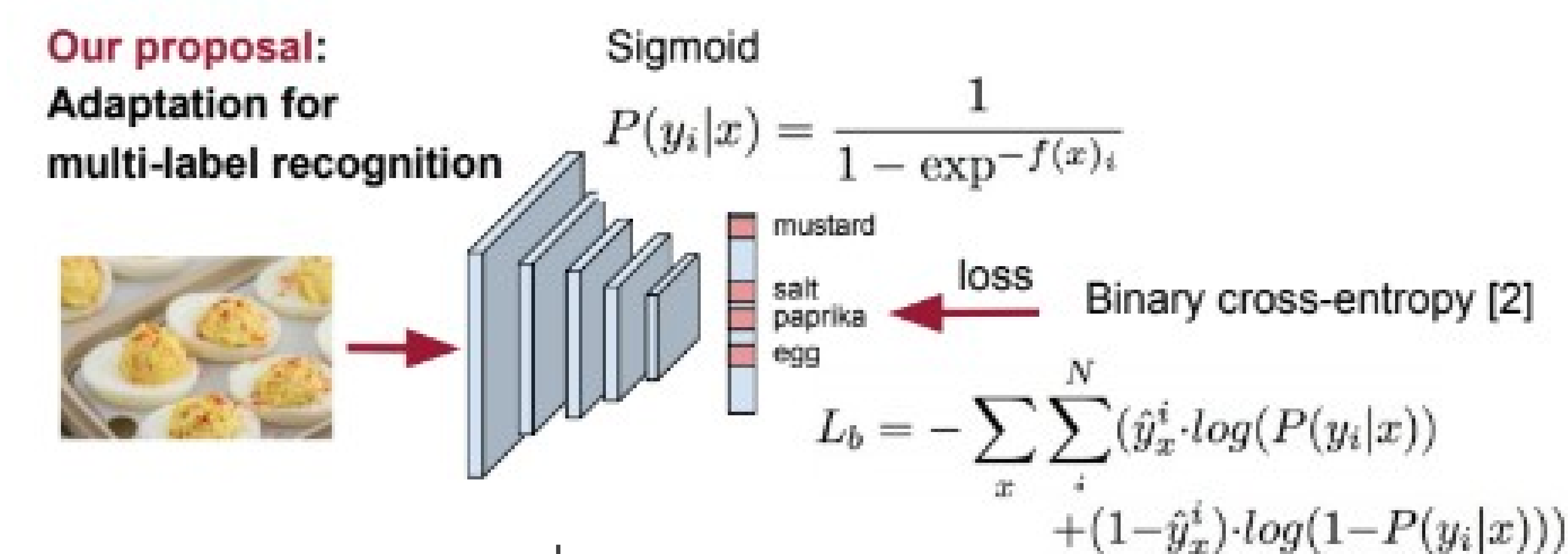
- Change the lost function to the Binary cross-entropy function $L_{b:}$





Multi-label classification

- Change the lost function to the Binary cross-entropy function L_b:

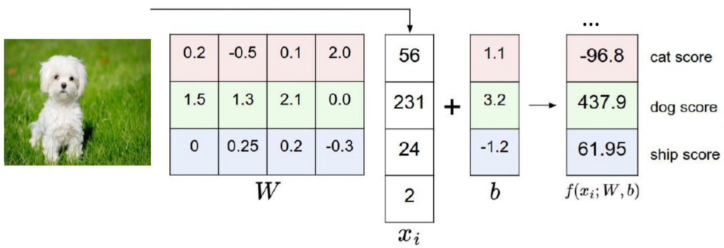


Index

- AI, Machine learning & Deep learning
- What is a Convolutional Neural Network?
  - Layers
  - Optimization
- Applications

How is a CNN doing deep learning?

$y=Wx$ $y_1=\Sigma_i W_{1i} x_i$ $y=W(Wx)$ $y=W(W(Wx))$



Fully connected layers

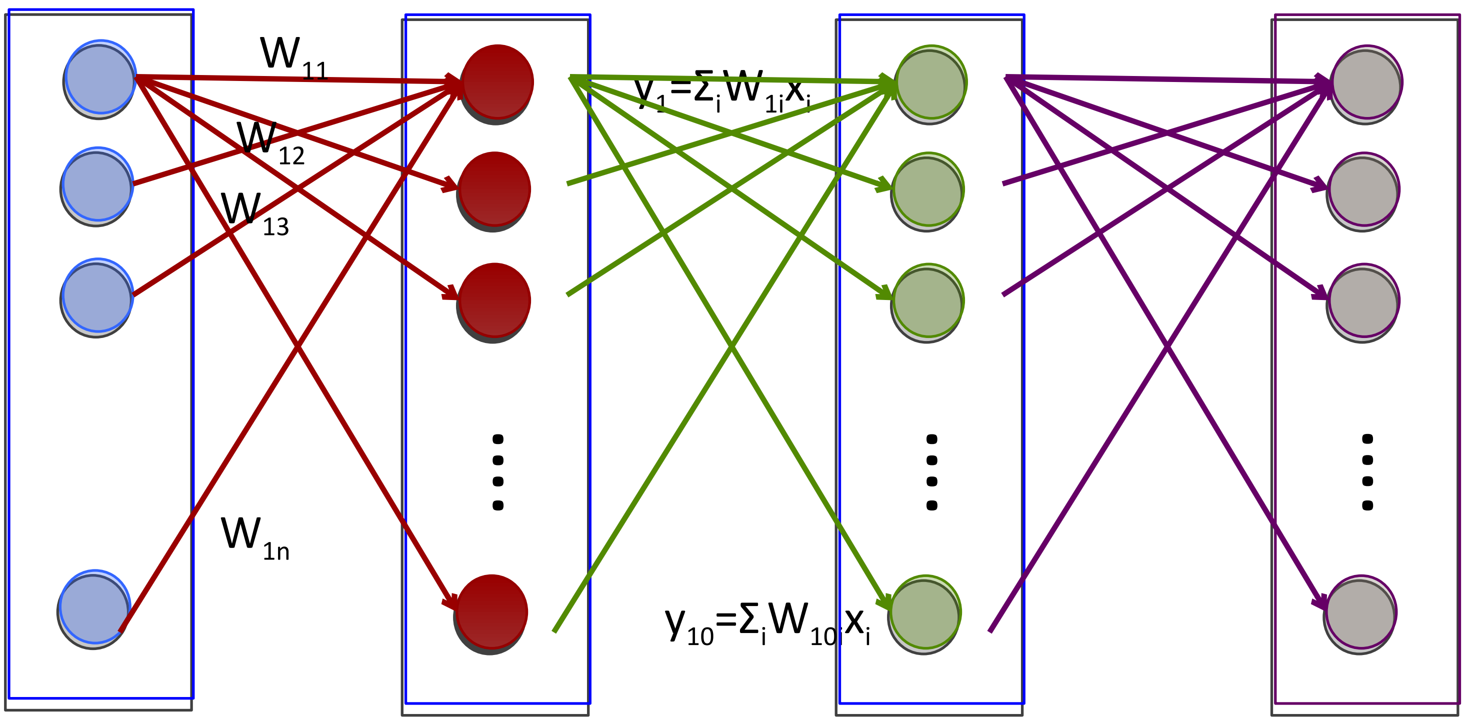
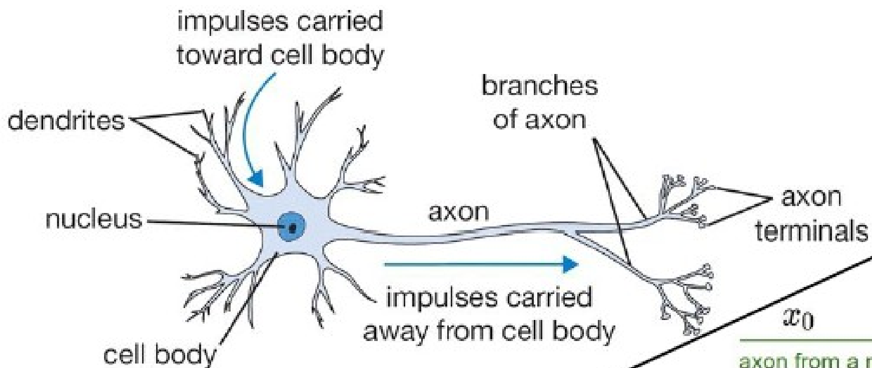
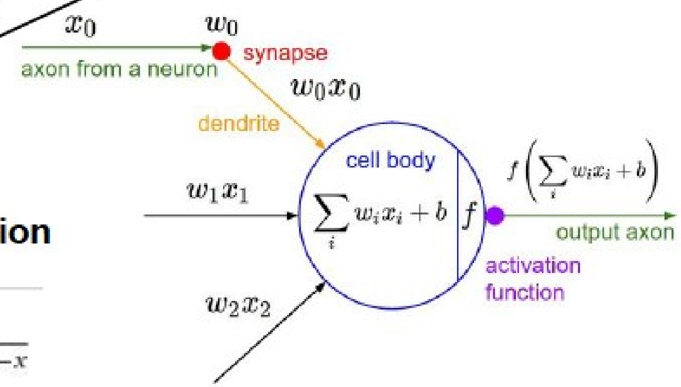
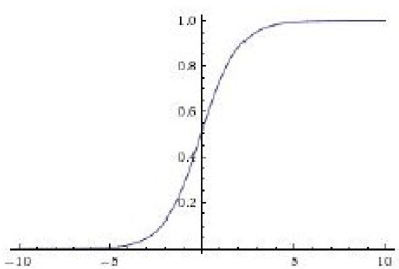


Image First layer Second layer Output layer

Why a CNN is a neural network?



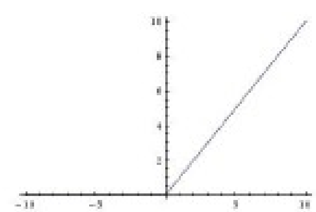




sigmoid activation function

$$\frac{1}{1+e^{-x}}$$

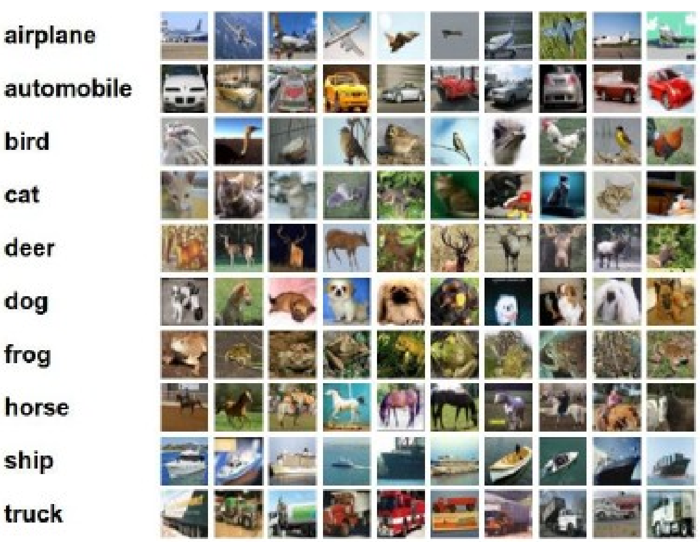
ReLU max(0,x)



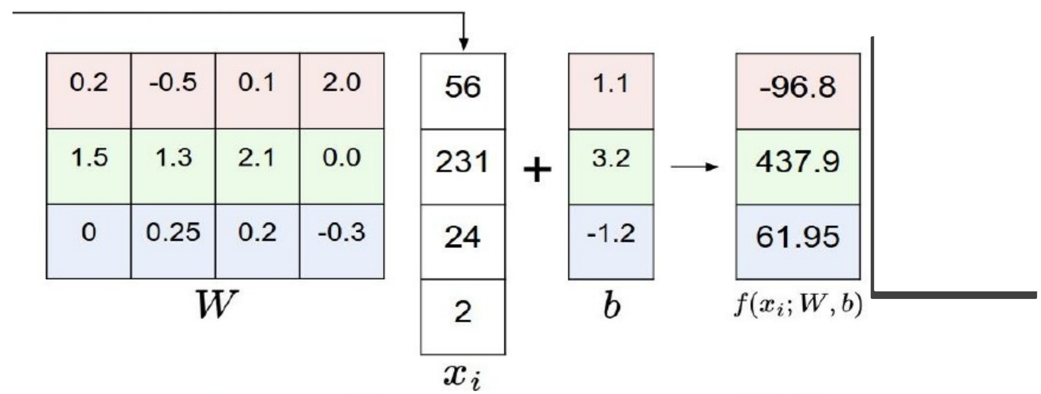
Modern CNNs – 10M neurons
Human CNNs – 5B of neurons.

From: Fei-Fei Li & Andrej Karpathy & Justin Johnson

Why is it convolutional?



$$f(x_i, W, b) = Wx_i + b$$

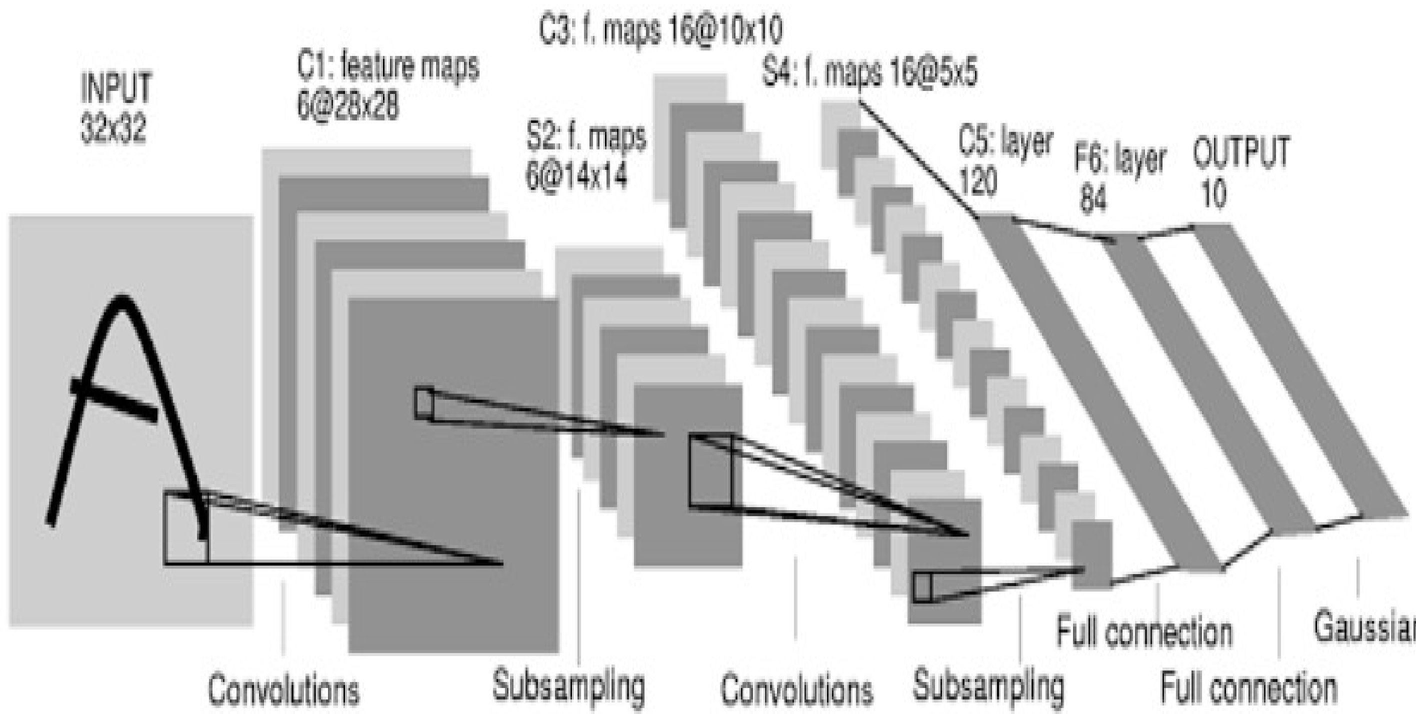


Adapted from: Fei-Fei Li & Andrej Karpathy & Justin Johnson

What is new in the Convolutional Neural Network?

1998

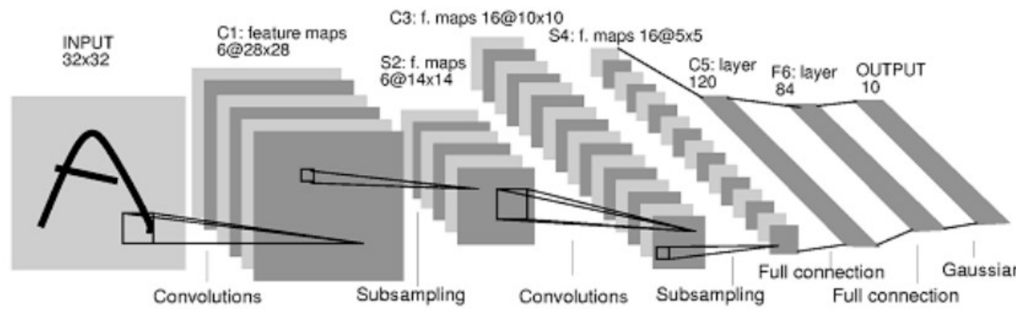
LeCun et al.



CNN evolution

1998

LeCun et al.



# of transistors

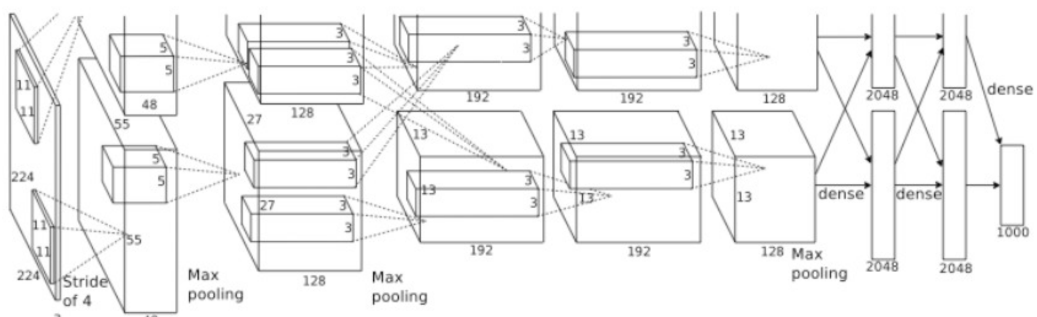
10^6

# of pixels used in training

10^7

2012

Krizhevsky et al.



# of transistors

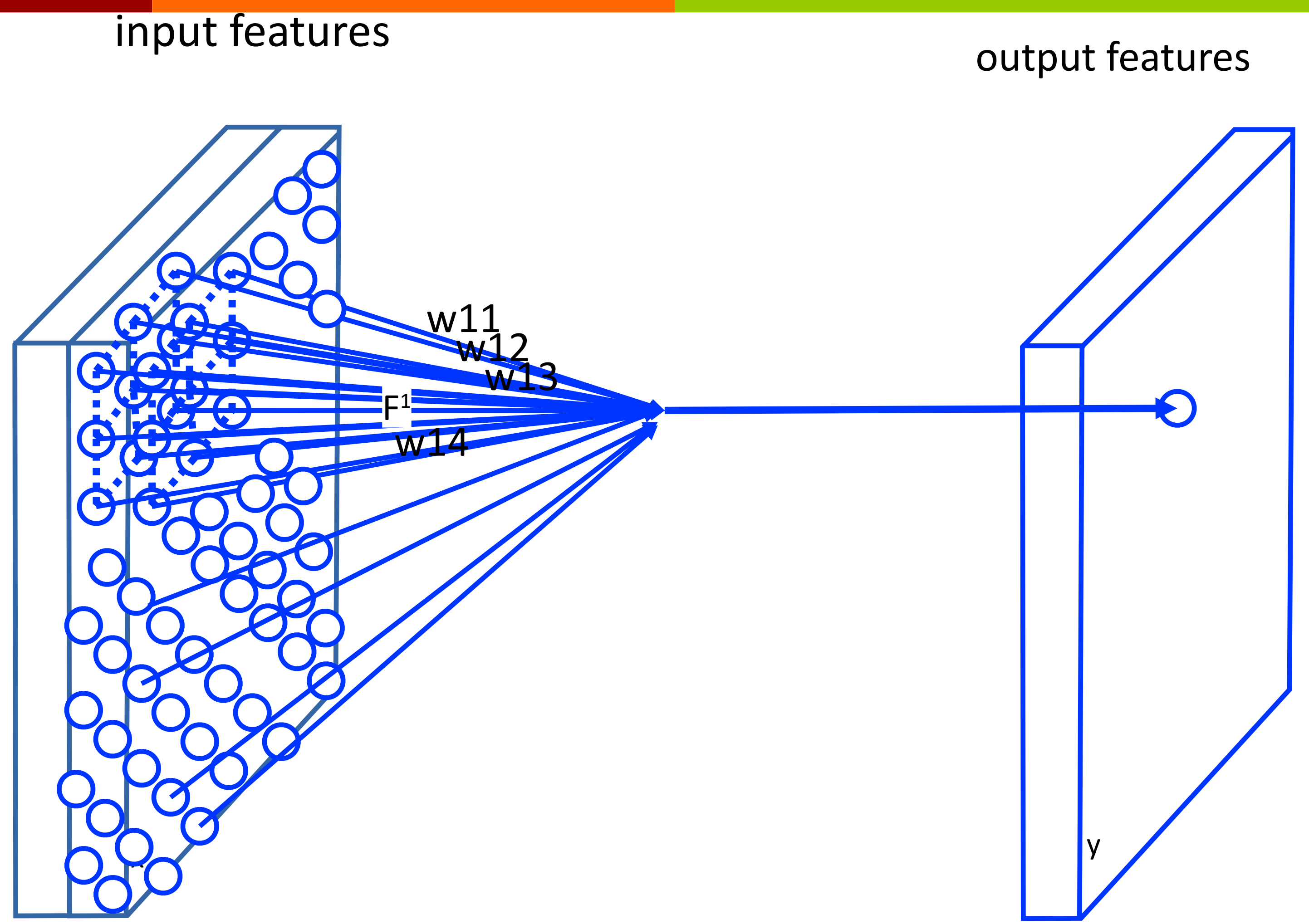
10^9

GPUs

# of pixels used in training

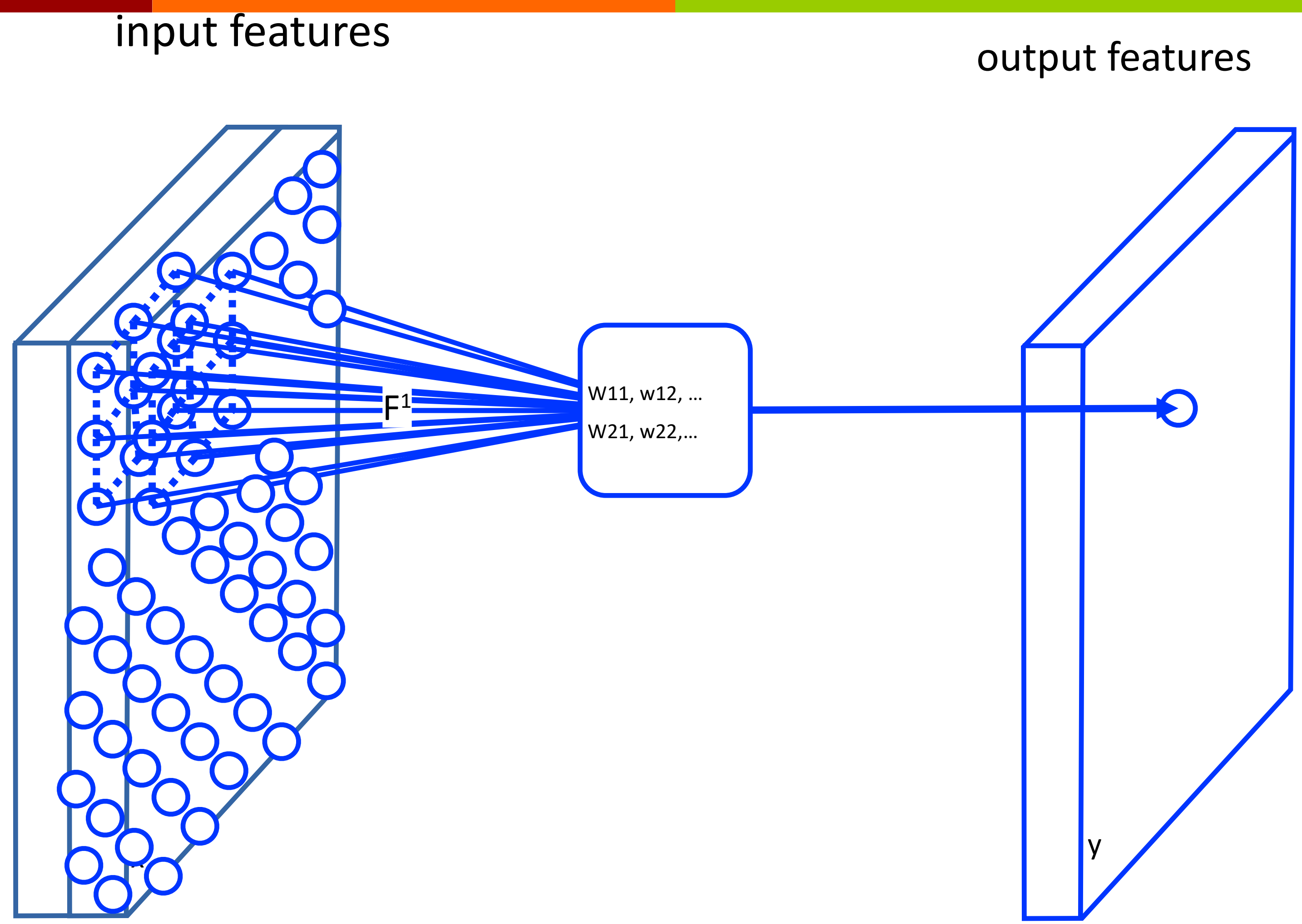
10^{14}

Why is it convolutional?



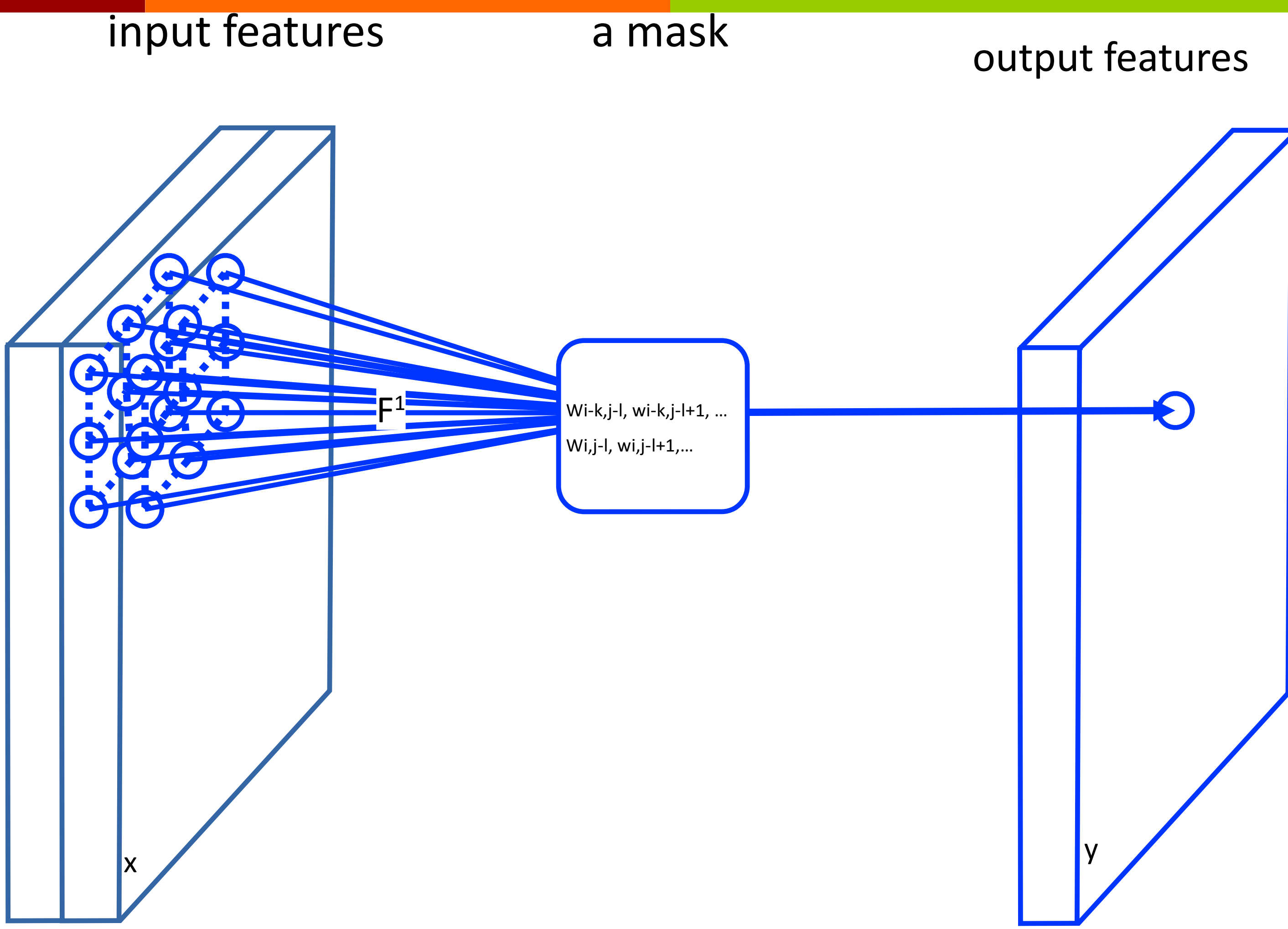
Let's consider a fully connected layer

Why is it convolutional?



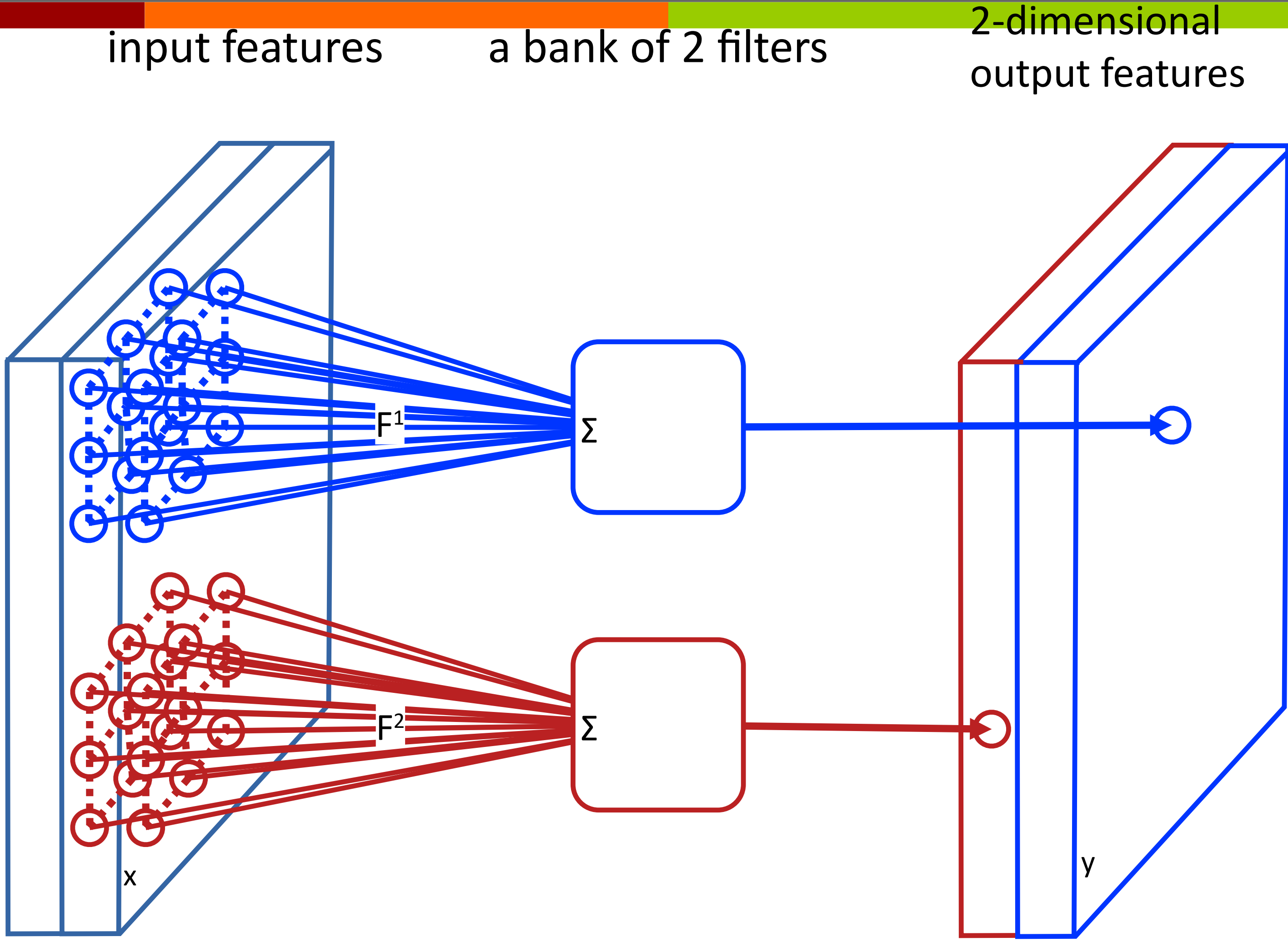
Let's consider a fully connected layer

Why is it convolutional?



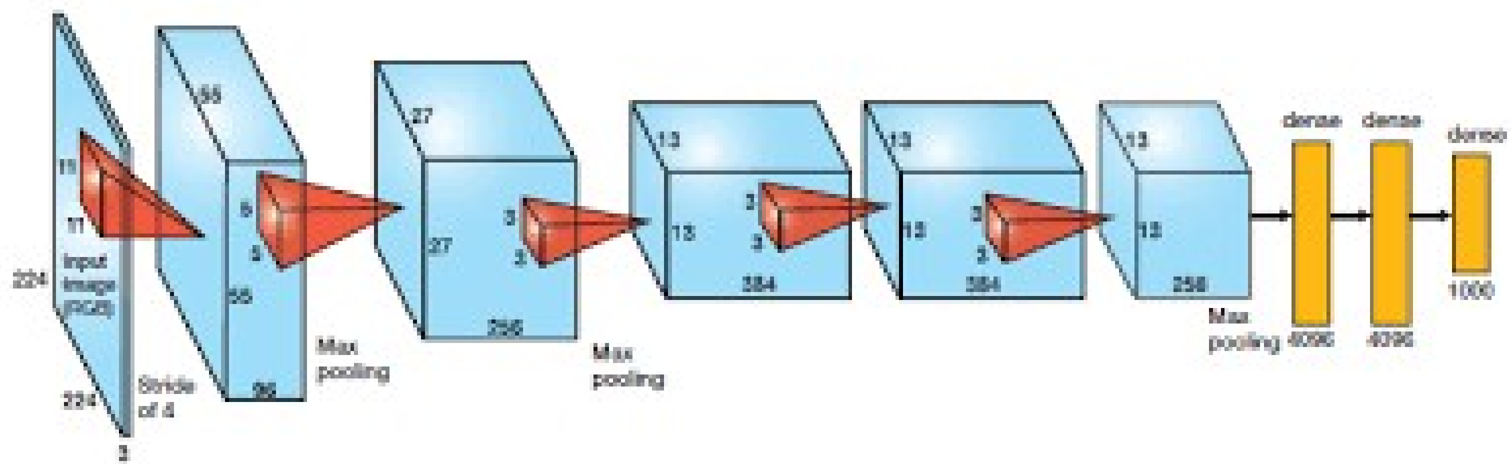
Convolution

Why is it convolutional?



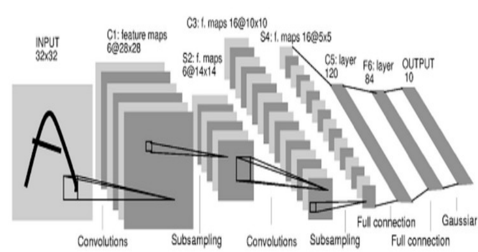
Convolution

Why is it convolutional?



1998

LeCun et al.

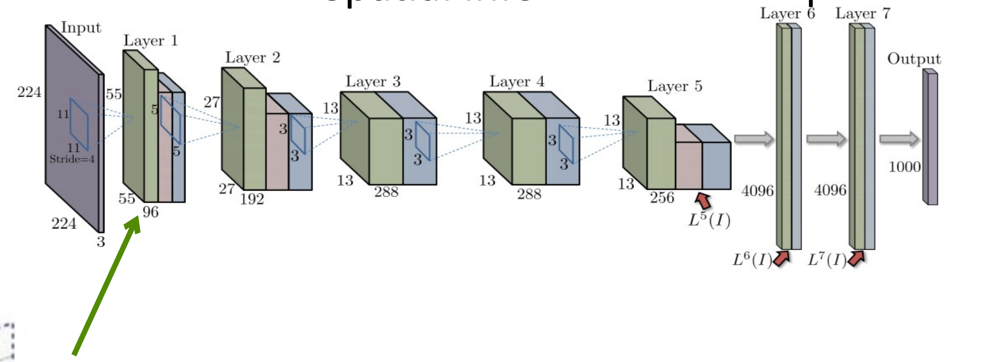


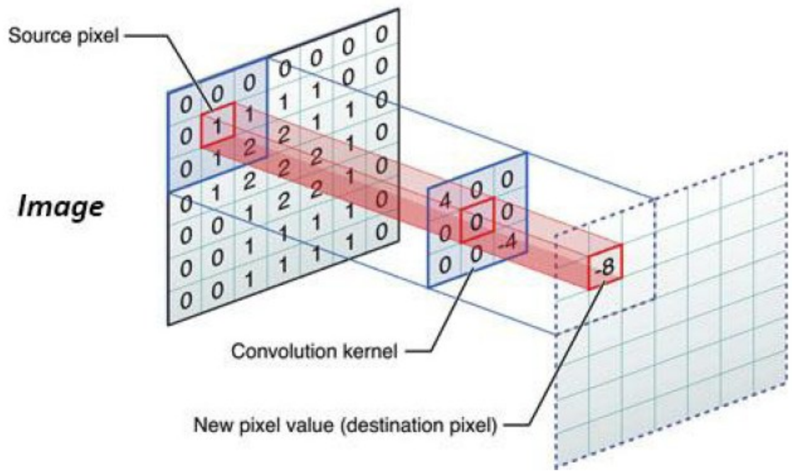
LeCun, Chief AI Scientist for **Facebook** AI Research (FAIR), and a Silver Professor at New York University

Convolutional layer

Spatial info

No spatial info





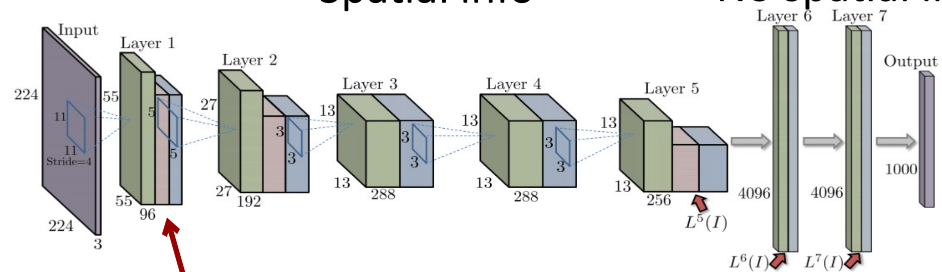
Filter Response

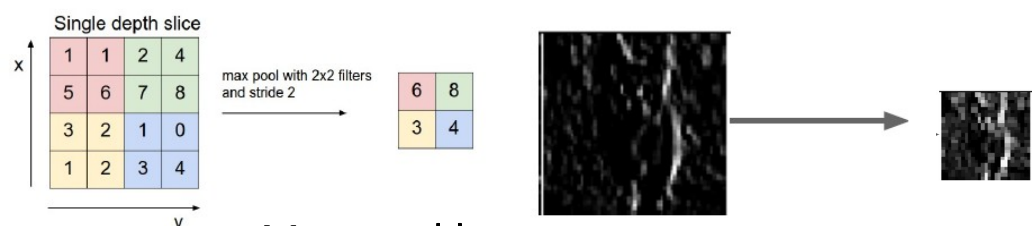
Convolutional layer

Max-pooling layer

Spatial info

No spatial info



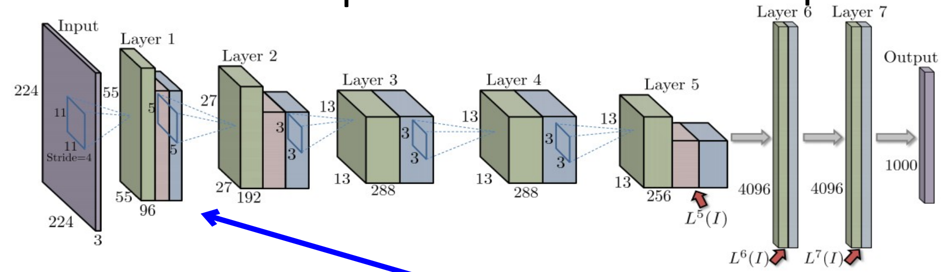


Max-pool layer

ReLu layer

Spatial info

No spatial info

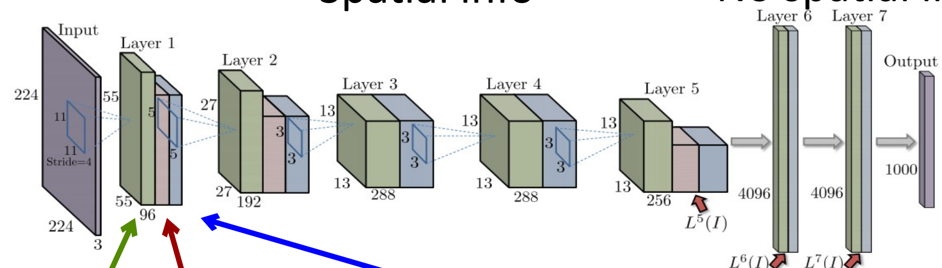


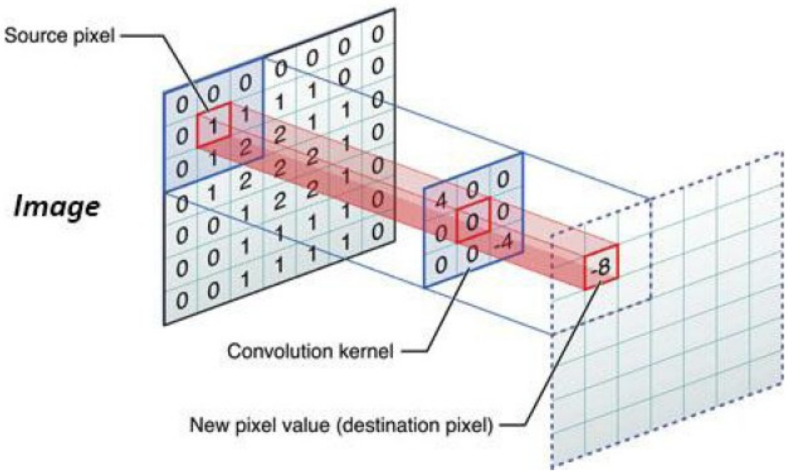
ReLU max(0,x)

All layers

Spatial info

No spatial info



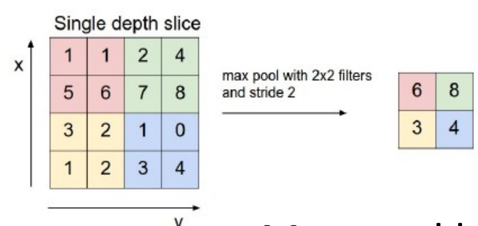


Image

Filter Response

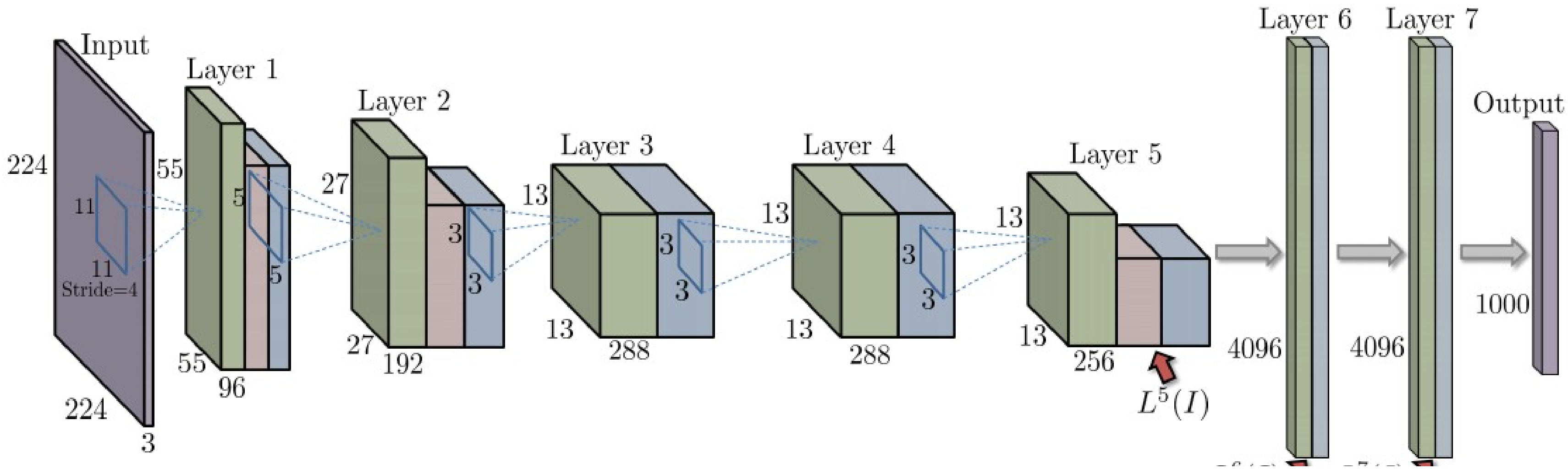
ReLU max(0,x)

Convolutional layer



Max-pool layer

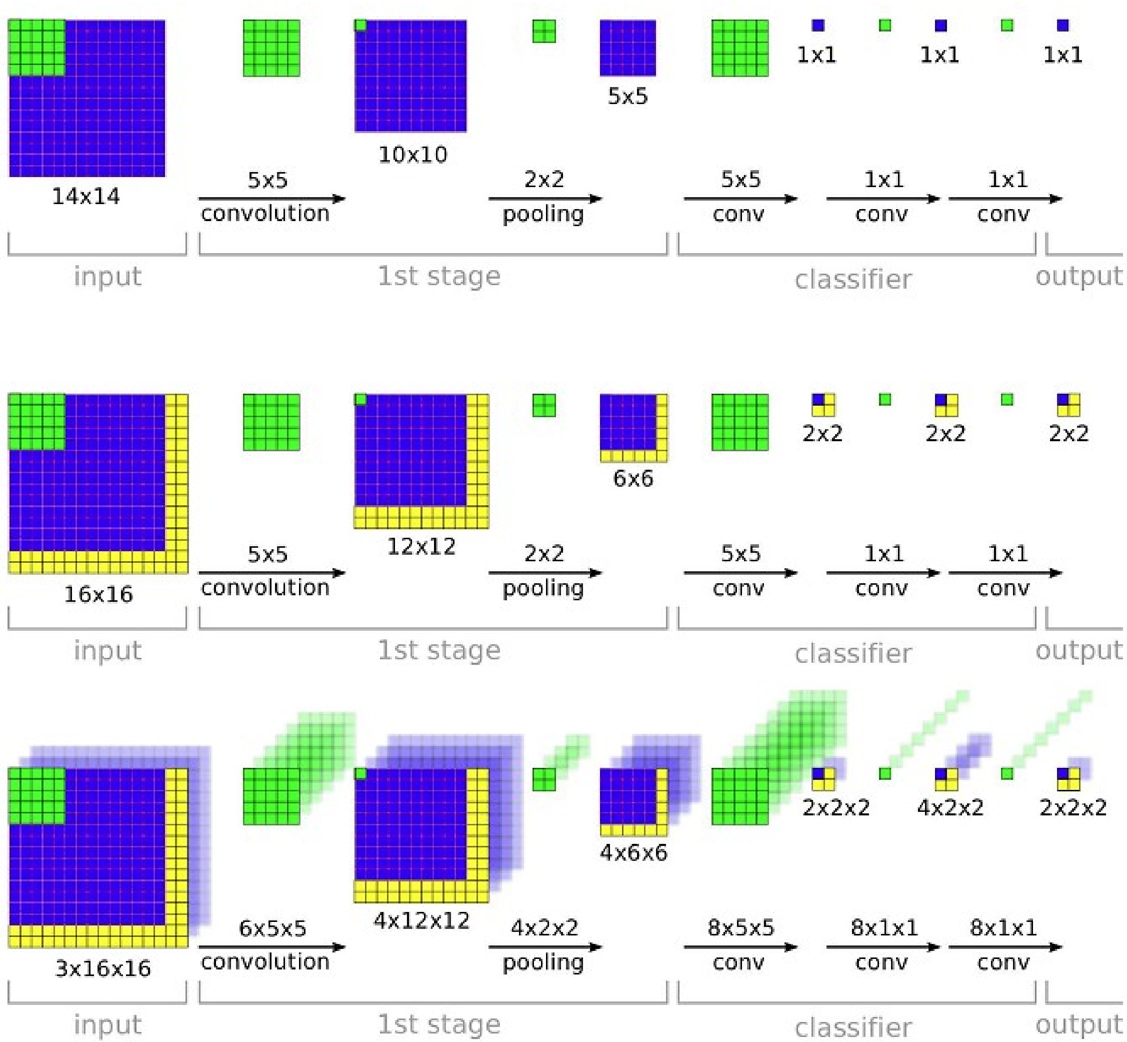
Training a CNN



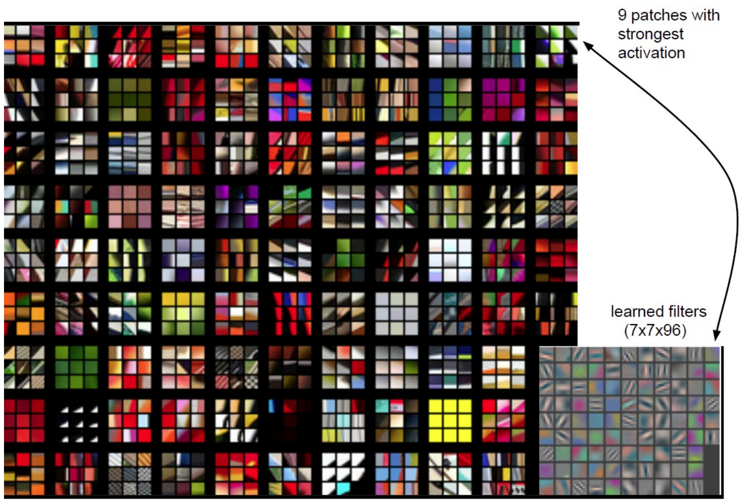
The process of training a CNN consists of training all hyperparameters: convolutional matrices and weights of the fully connected layers.

- **Several millions of parameters!!!**

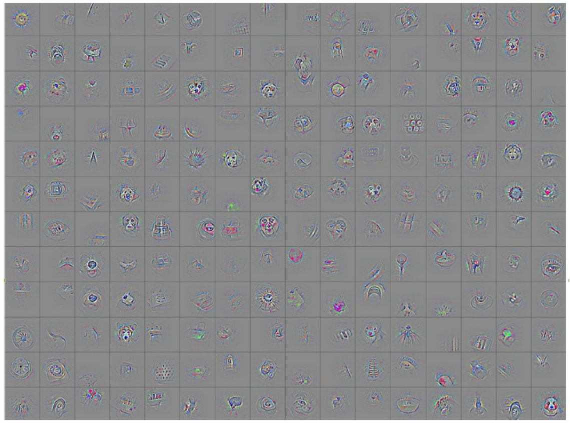
How does the CNN work?



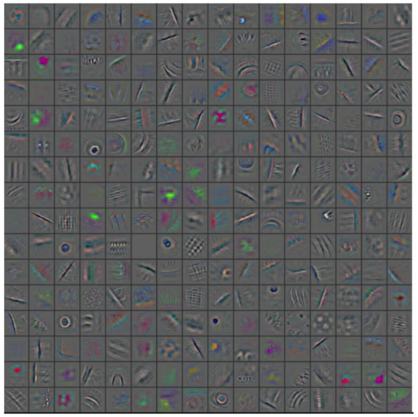
Learned convolutional filters



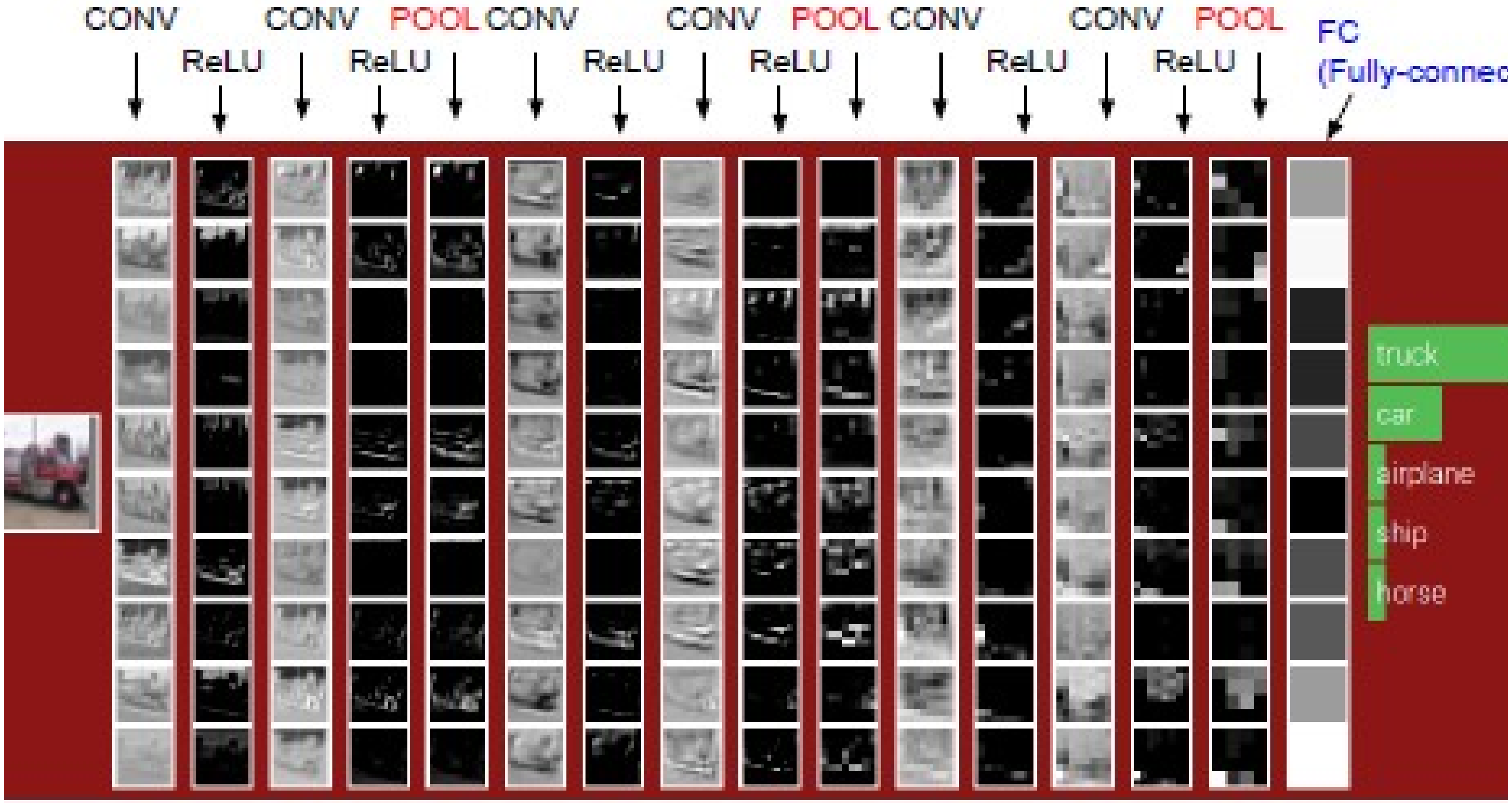
Strongest activations: Stage 5



Strongest activations: Stage 2



Example architecture



The trick is to train the weights such that when the network sees a picture of a truck, the last layer will say "truck".

Credit slide: Li Fei-fei

1001 benefits of CNN

Transfer learning: Fine tunning for object recognition

Replace and retrain the classier on top of the ConvNet

Fine-tune the weights of the pre-trained network by continuing the backpropagation

Feature extraction by CNN

Object detection

Object segmentation

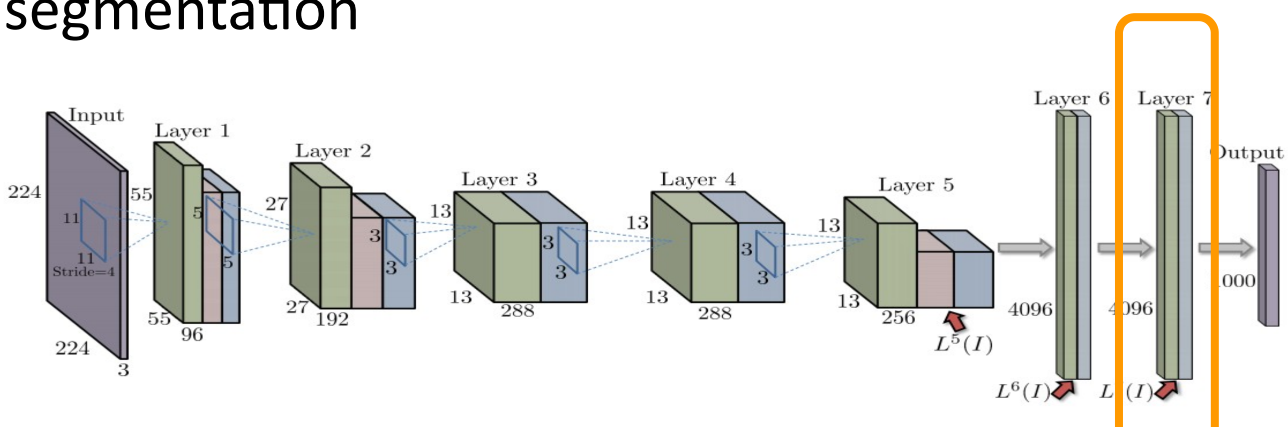


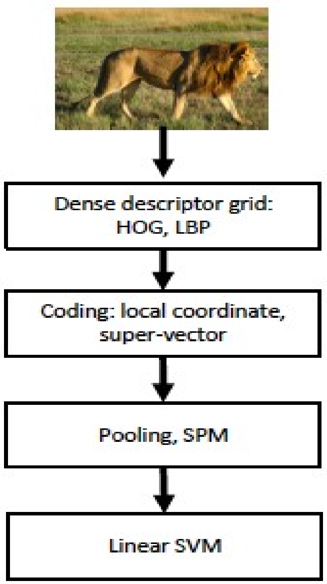
Image similarity and matching by CNN

Convolutional Neural Networks (4096 Features)

IMAGENET Large Scale Visual Recognition Challenge

Year 2010

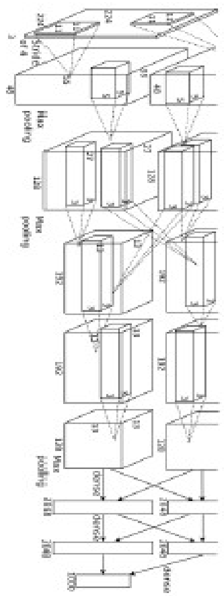
NEC-UIUC



[Lin CVPR 2011]

Year 2012

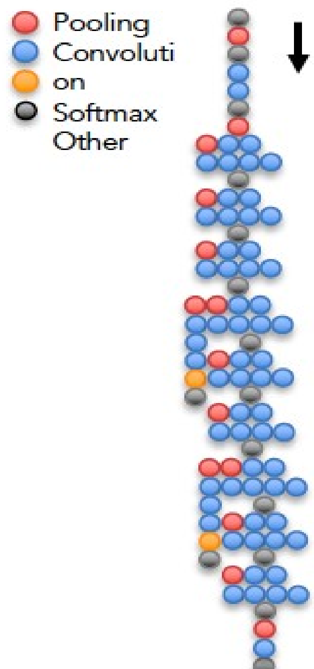
SuperVision



[Krizhevsky NIPS 2012]

Year 2014

GoogLeNet



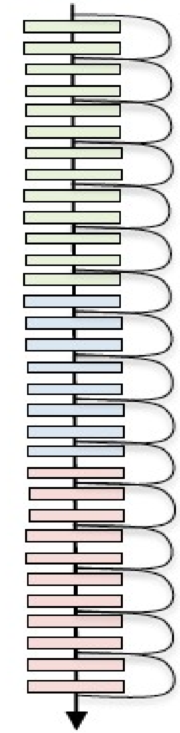
[Szegedy arxiv 2014]

VGG

[Simonyan arxiv 2014]

Year 2015

MSRA

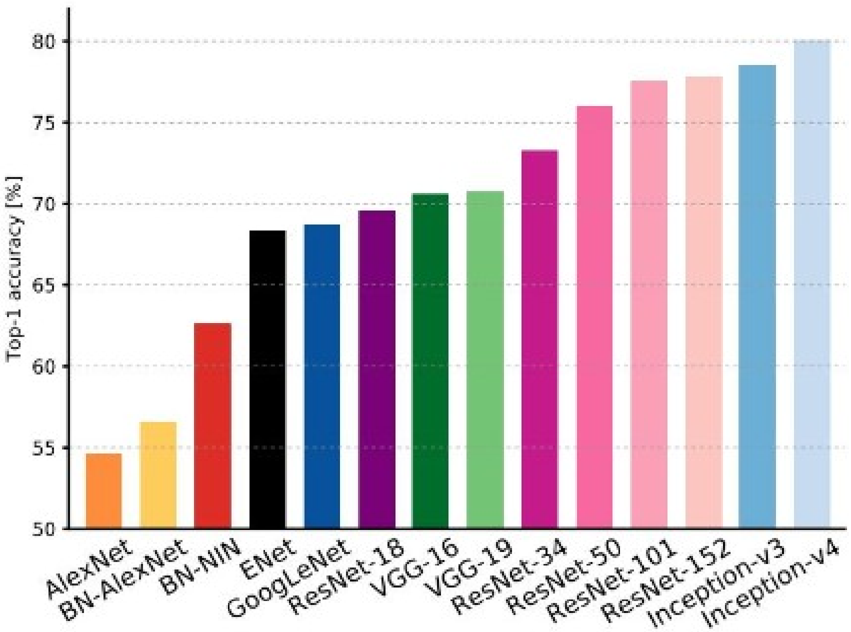


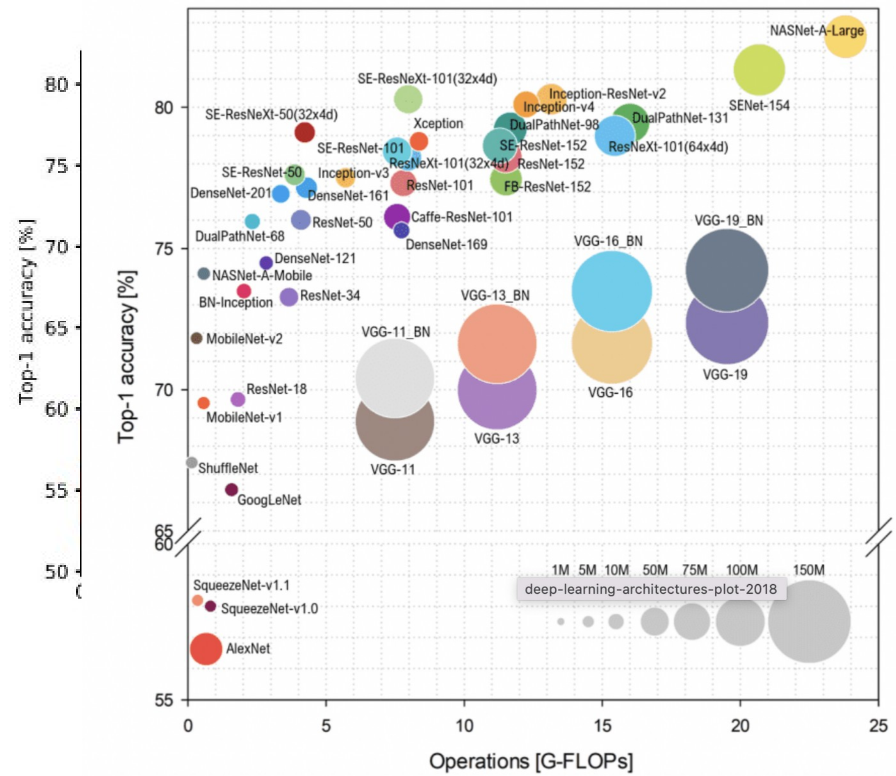
[He ICCV 2015]

))))

14:40

Analysis of CNNs





- **Millions of parameters!!!**

The process of training a CNN consists of training all hyperparameters: convolutional matrices and weights of the fully connected layers.

Index

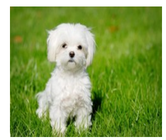
- AI, Machine learning & Deep learning
- What is a Convolutional Neural Network?
  - Layers
  - Loss function and CNN Optimization
- Applications

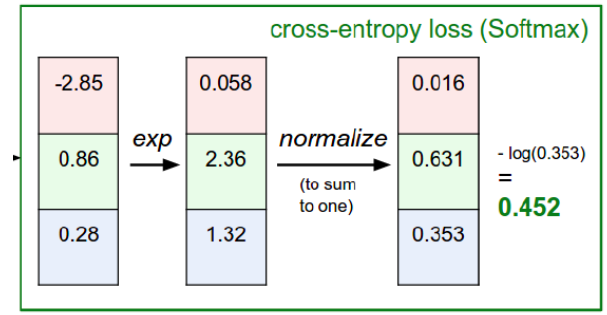
Loss function and optimisation

- **Question:** if you were to assign a single number to how unhappy you are with these scores, what would you do?

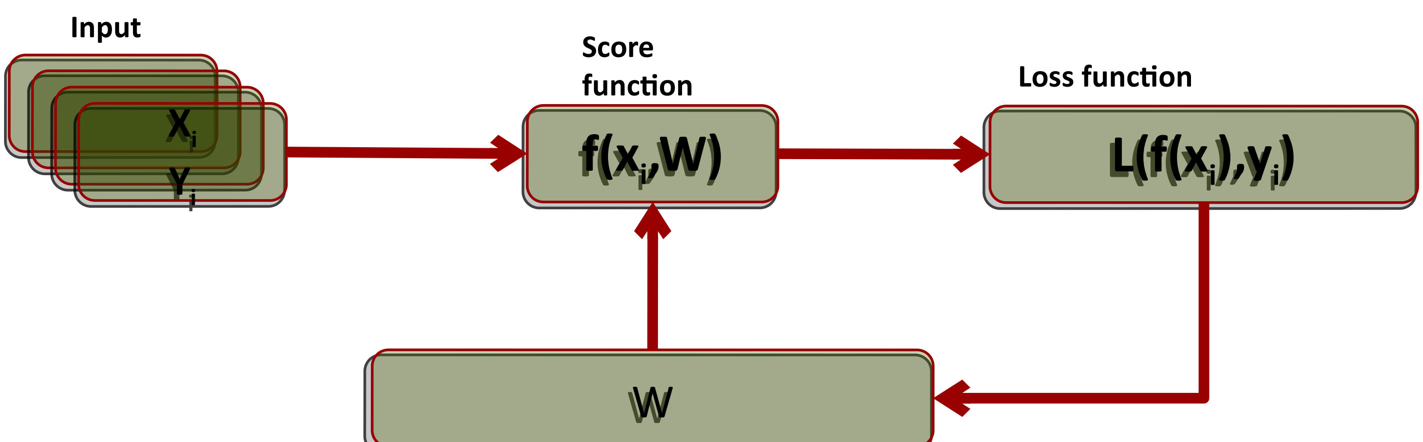
$$L_i = -\log\left(\boxed{\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}}\right)$$

softmax function

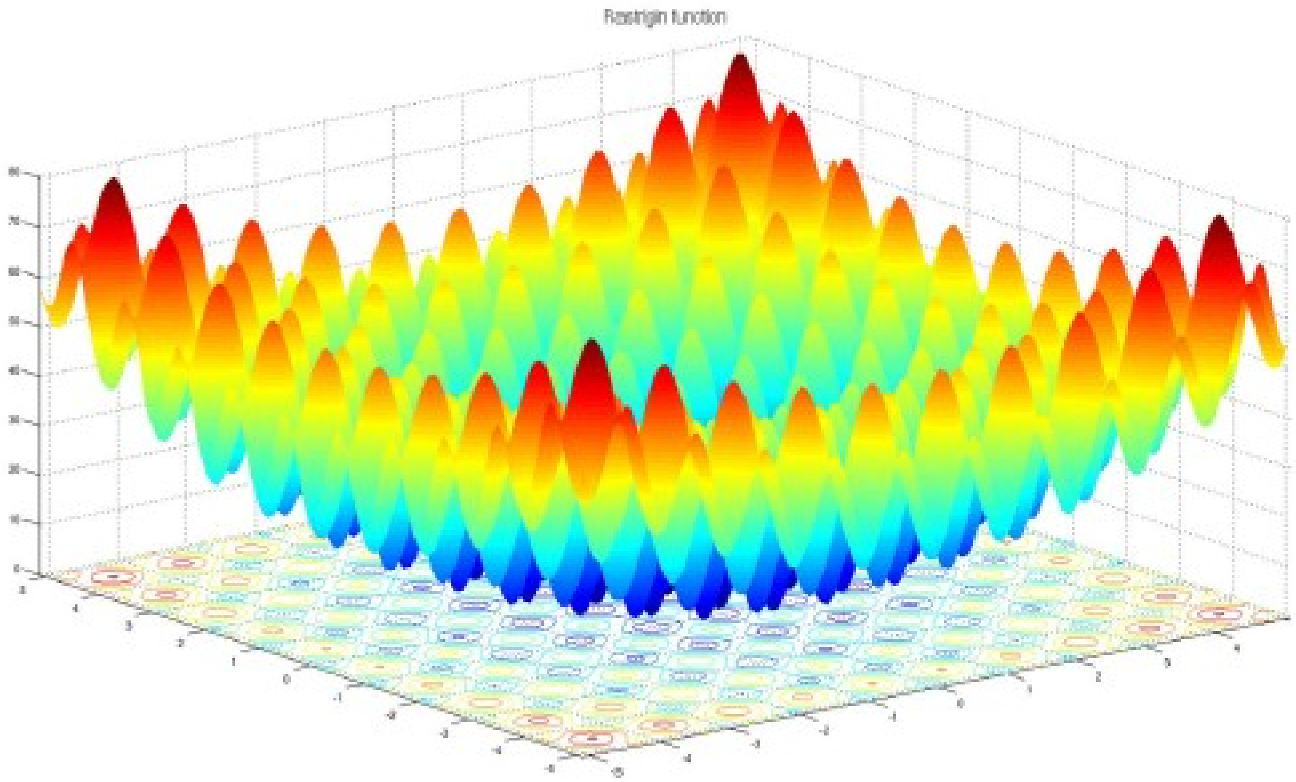




Question : Given the score and the loss function, how to find the parameters W?



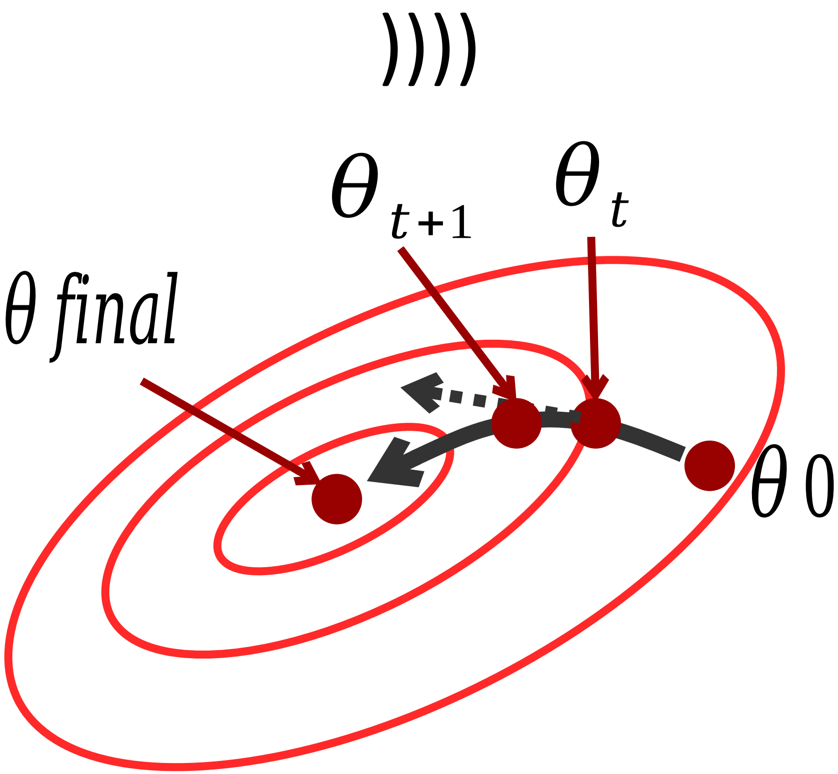
Optimization



The process of training a CNN consists of training all hyperparameters: convolutional matrices and weights of the fully connected layers.

- **Millions of parameters!!!**

Gradient descent



- Initialize randomly
- For t in 0,…, $T_{maxiter}$

Gradient of the objective

Learning rate (step size)

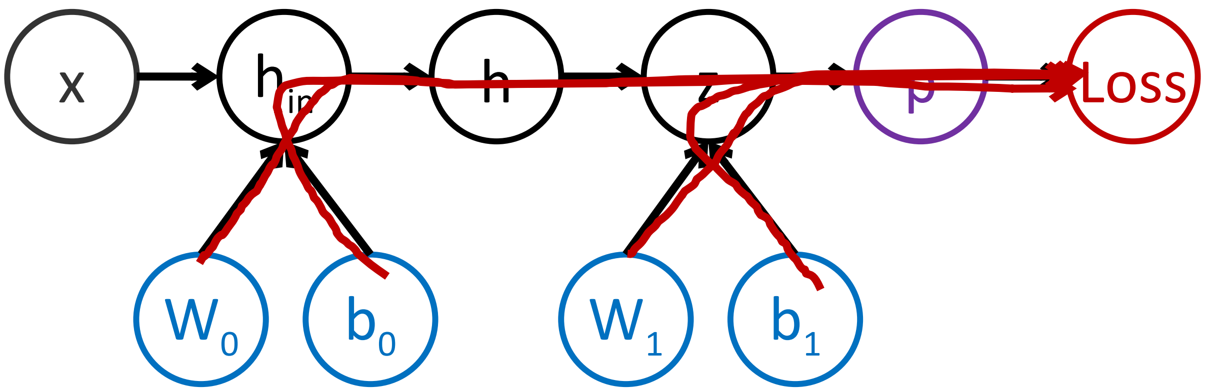
- Computation of requires a full sweep over the training data
- Per-iteration comp. cost = O(n)

Chain rule

- Identify how each variable influence the loss

$$\frac{\partial \mathrm{Loss}}{\partial W_1} = \frac{\partial Loss}{\partial p} \cdot \ldots \cdot \frac{\partial z}{\partial W_1}$$

$$\frac{\partial \mathrm{Loss}}{\partial b_1} = \frac{\partial Loss}{\partial p} \cdot \ldots \cdot \frac{\partial z}{\partial b_1}$$

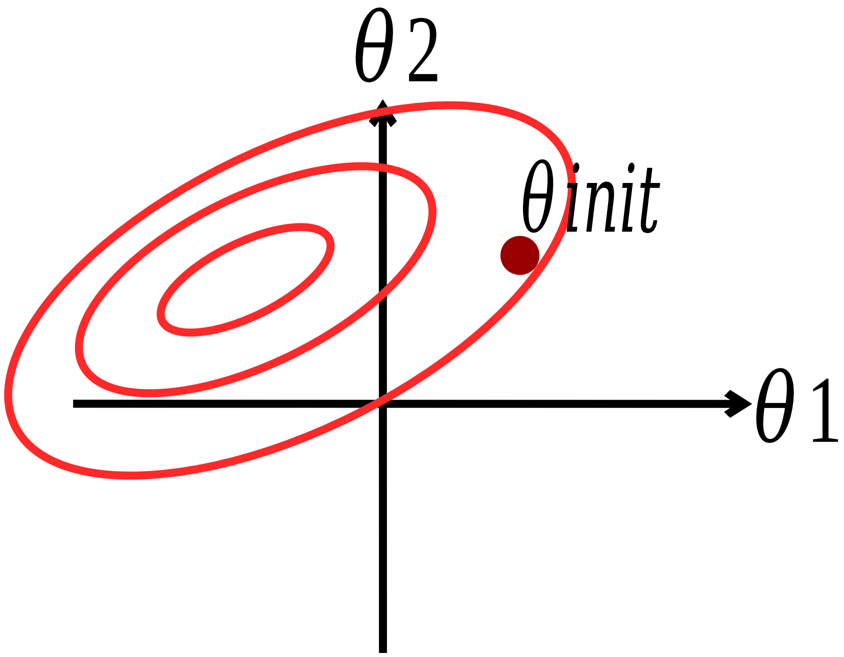


$$\frac{\partial \mathrm{Loss}}{\partial W_0} = \frac{\partial Loss}{\partial p} \cdot \ldots \cdot \frac{\partial h_{in}}{\partial W_0}$$

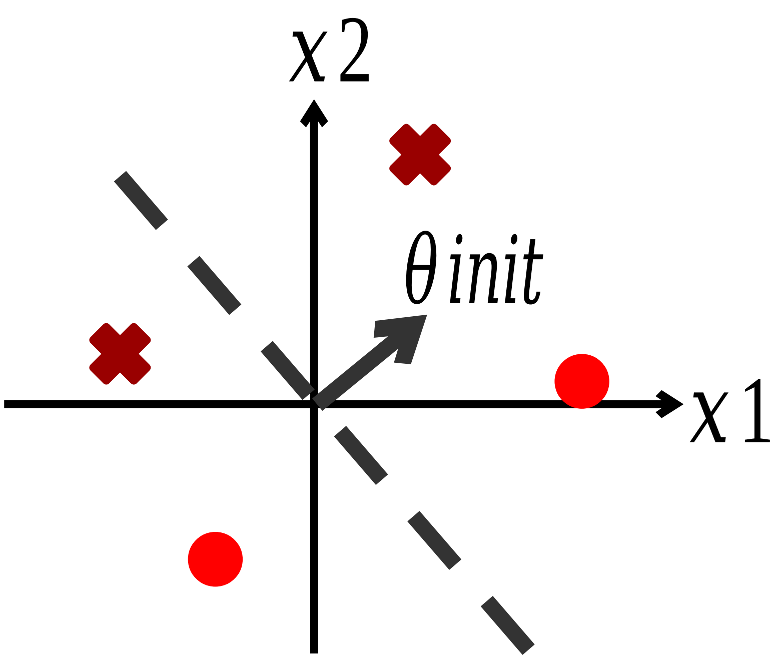
$$\frac{\partial \mathrm{Loss}}{\partial b_0} = \frac{\partial Loss}{\partial p} \cdot \ldots \cdot \frac{\partial h_{in}}{\partial b_0}$$

Landscape of training objective

Parameter space

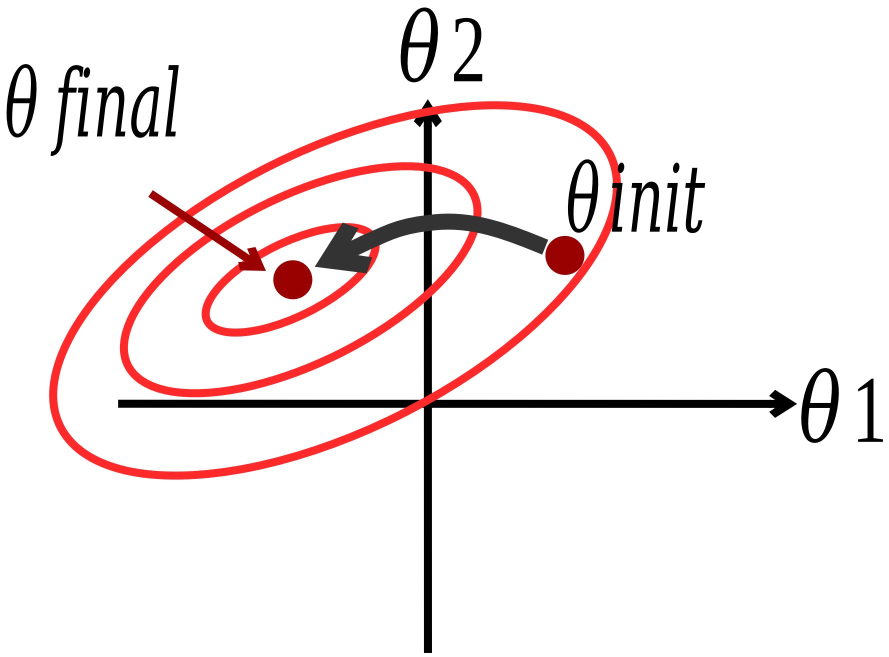


Example space

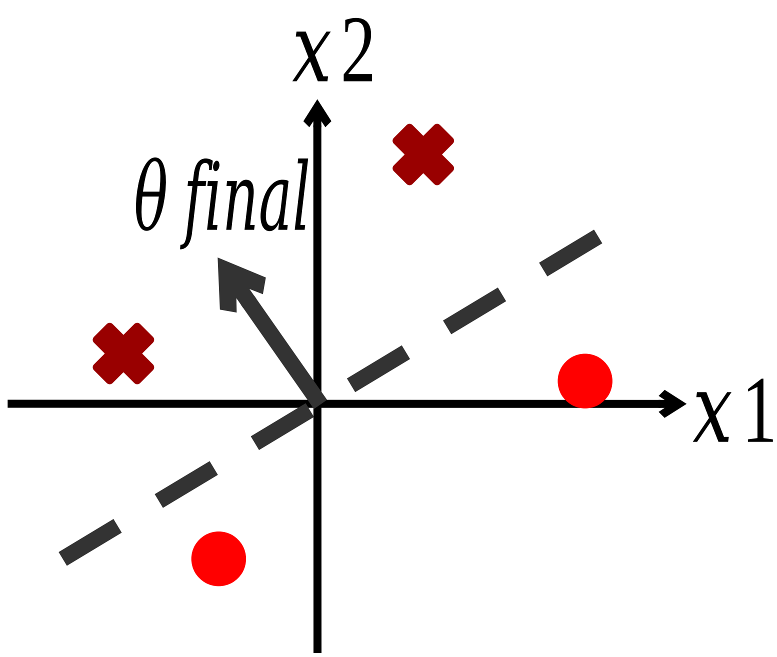


Landscape of training objective

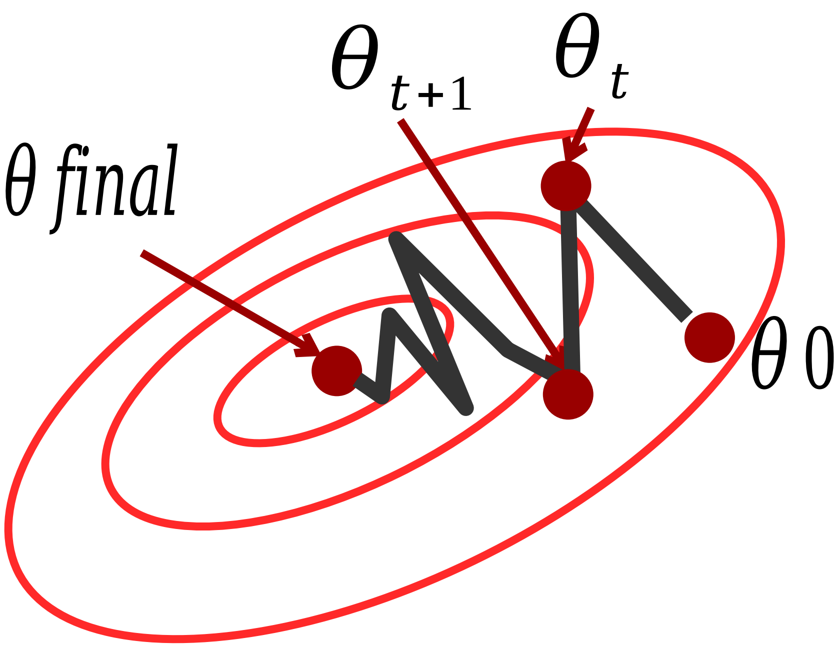
Parameter space



Example space



Stochastic gradient descent (SGD)



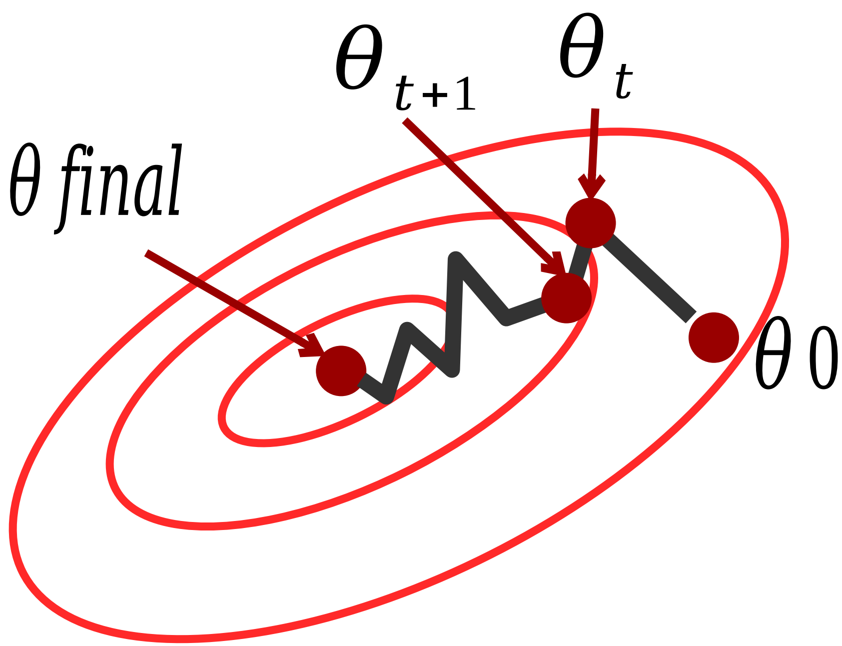
- Initialize randomly
- For t in 0,…, $T_{maxiter}$

Stochastic gradient

where index i is chosen randomly

- computation of requires only one training example
- Per-iteration comp. cost = O(1)

Mini-batch stochastic gradient descent



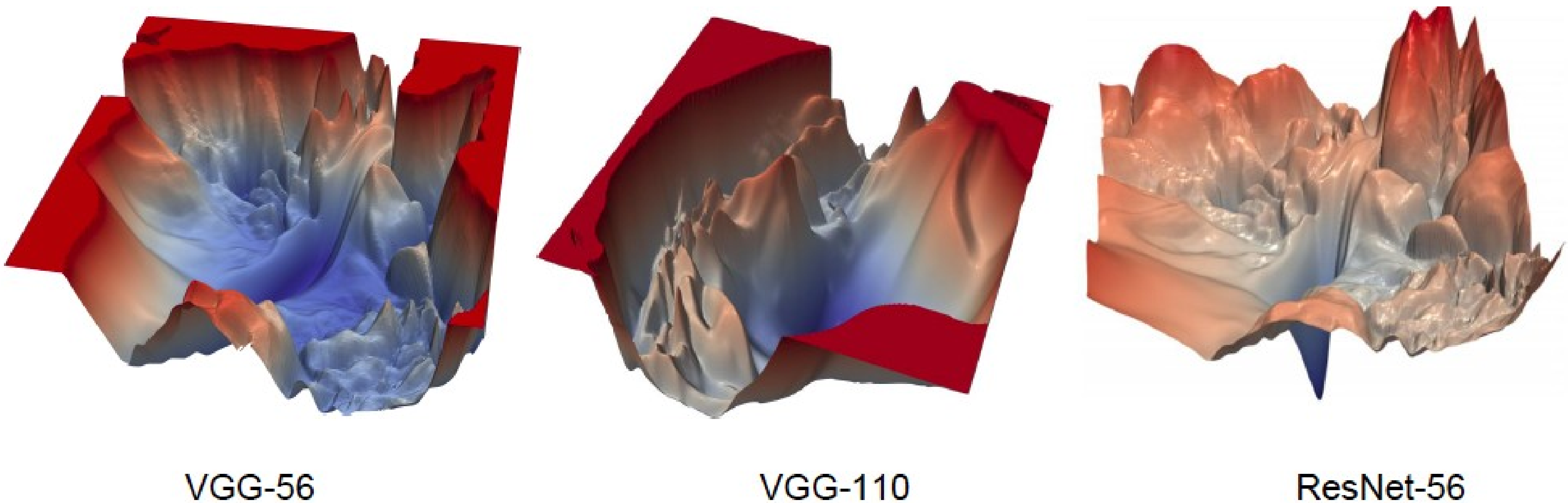
- Initialize randomly
- For t in 0,…, T_{maxiter}

$\underbrace{\qquad\qquad}_{\text{minibatch gradient}}$

where minibatch B is chosen randomly

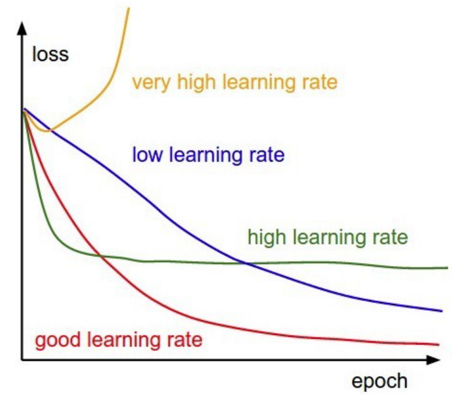
- is average gradient over random subset of data of size B
- Per-iteration comp. cost = O(B)

Stochastic Gradient Descent

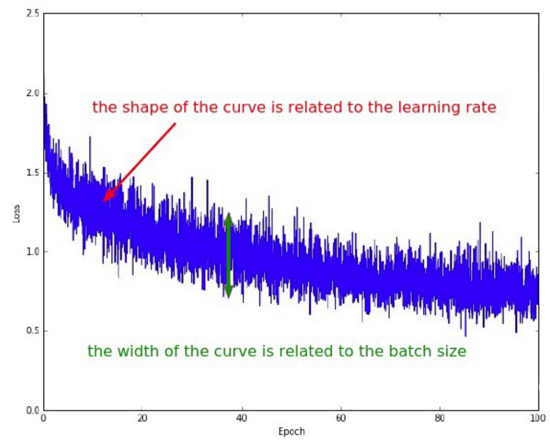


Hao Li et al., NIPS, 2017

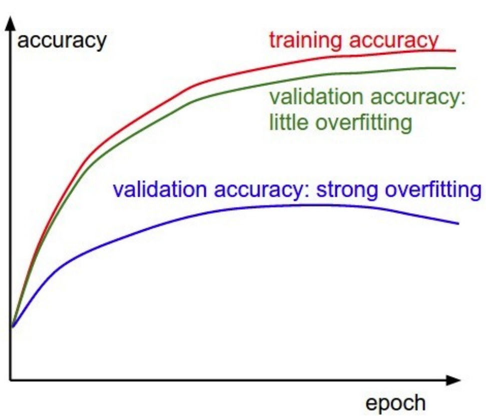
Monitoring loss and accuracy

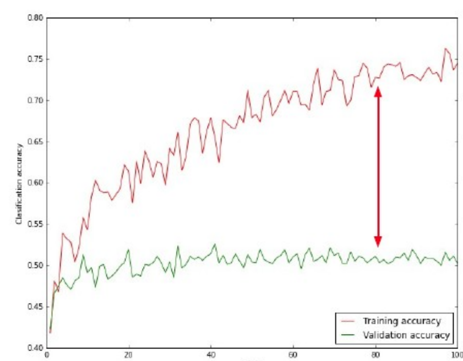


Adapted from Fei Fei slides



Looks linear?
Learning rate too low!
Decreases too slowly?
Learning rate too high.
Looks too noisy?
Increases the batch size.

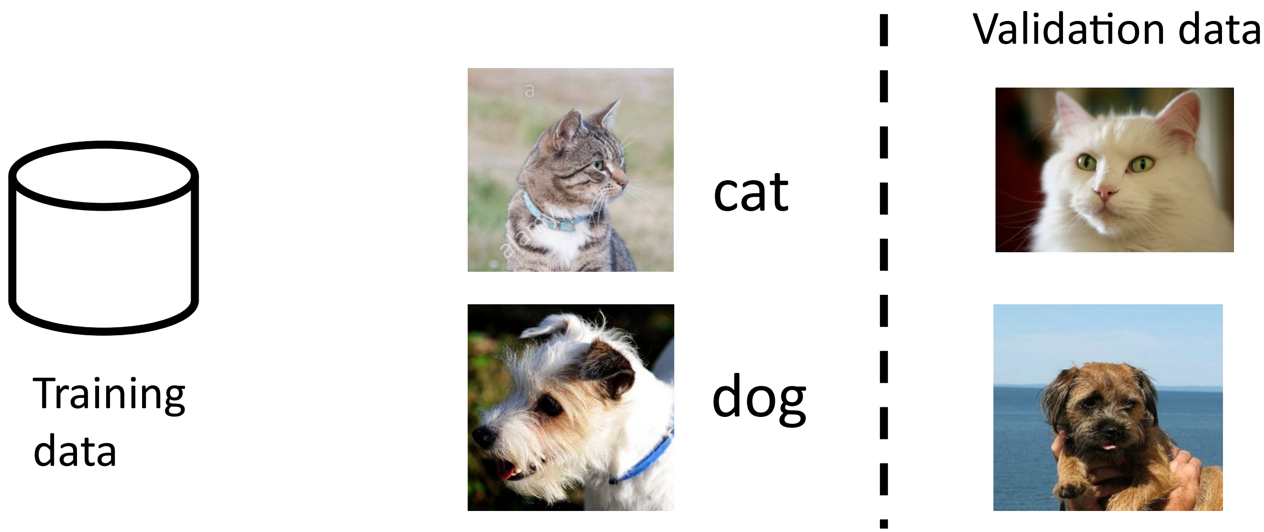




Big gap?
- you're overfitting,
increase
regularization!

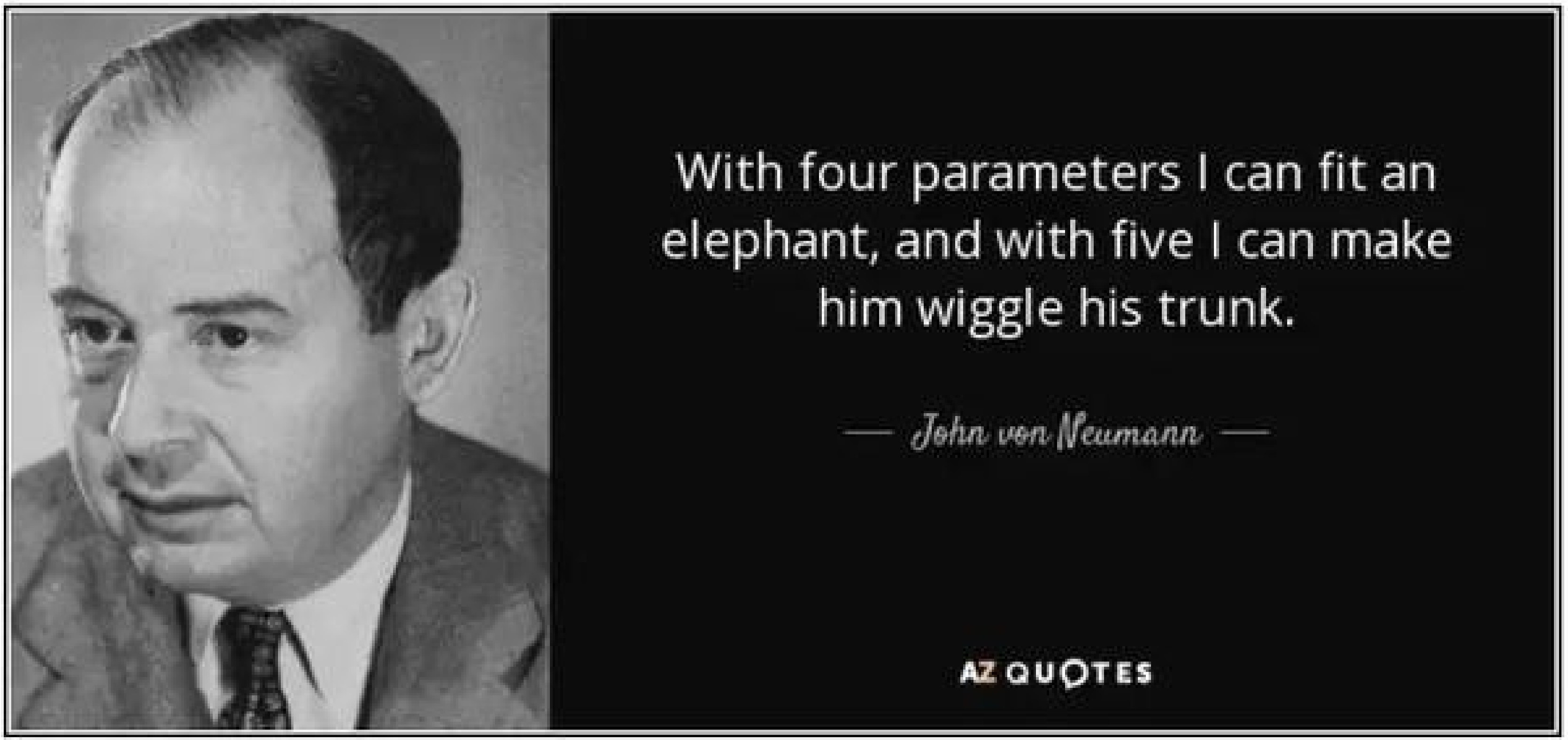
Overfitting – what is signal vs noise?

- Imagine:

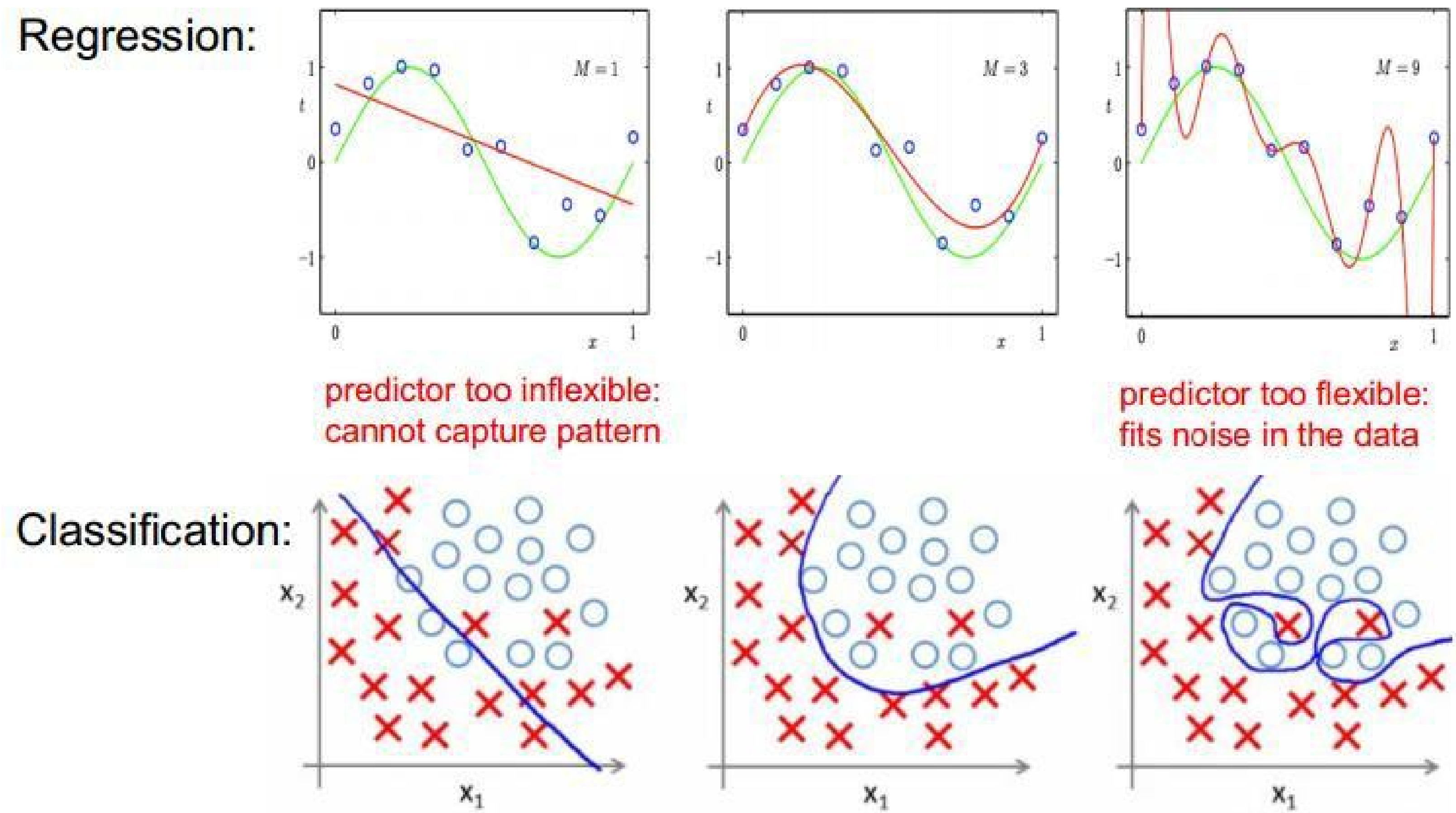


- Powerful models are more likely to overfit
- We need validation data: leave out some portion of the training data to validate the generalizability of the model

Overfitting

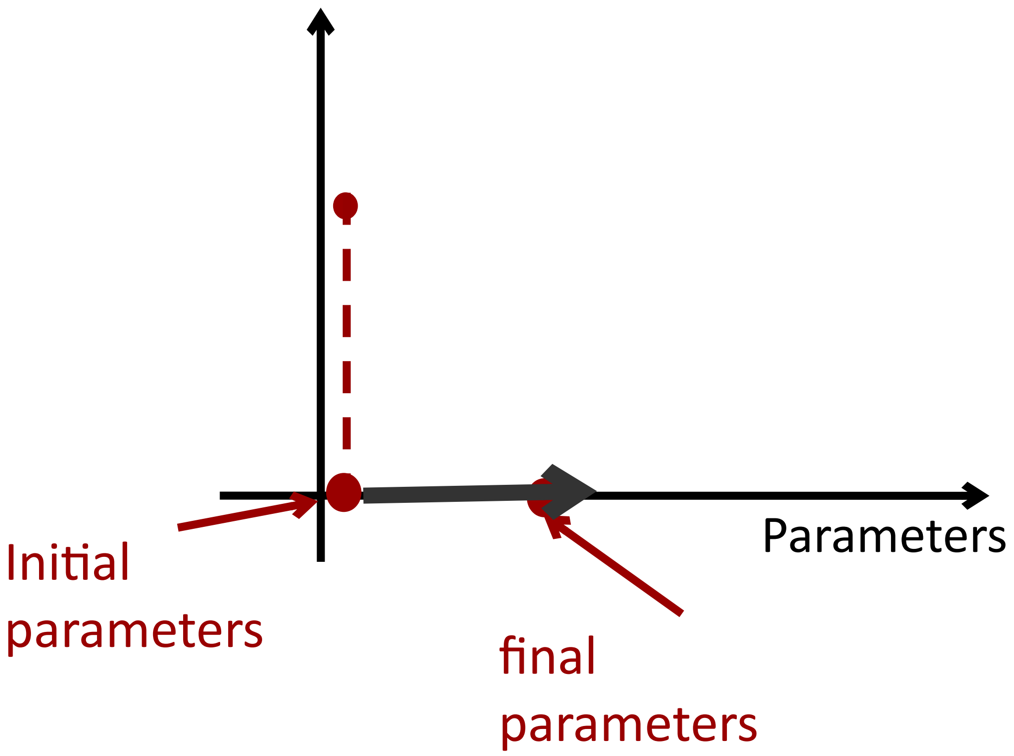


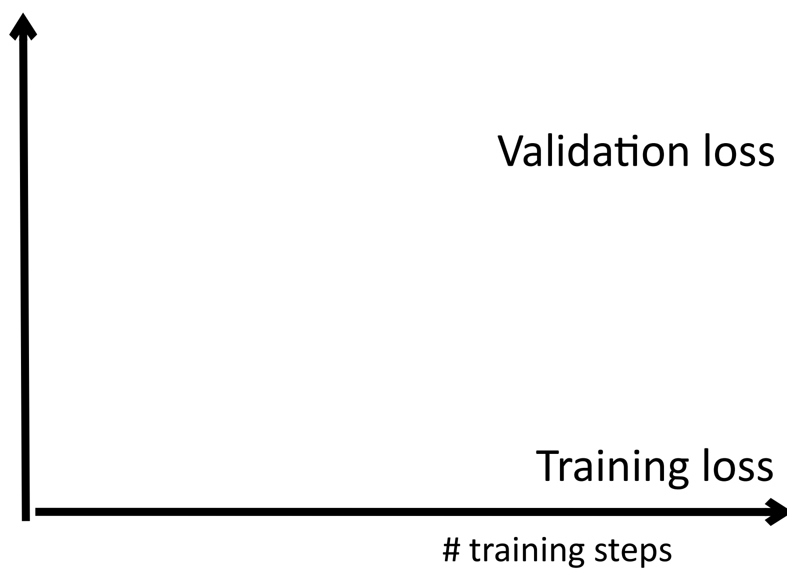
Under- and Over-fitting



Landscape of training objective

Objective function:



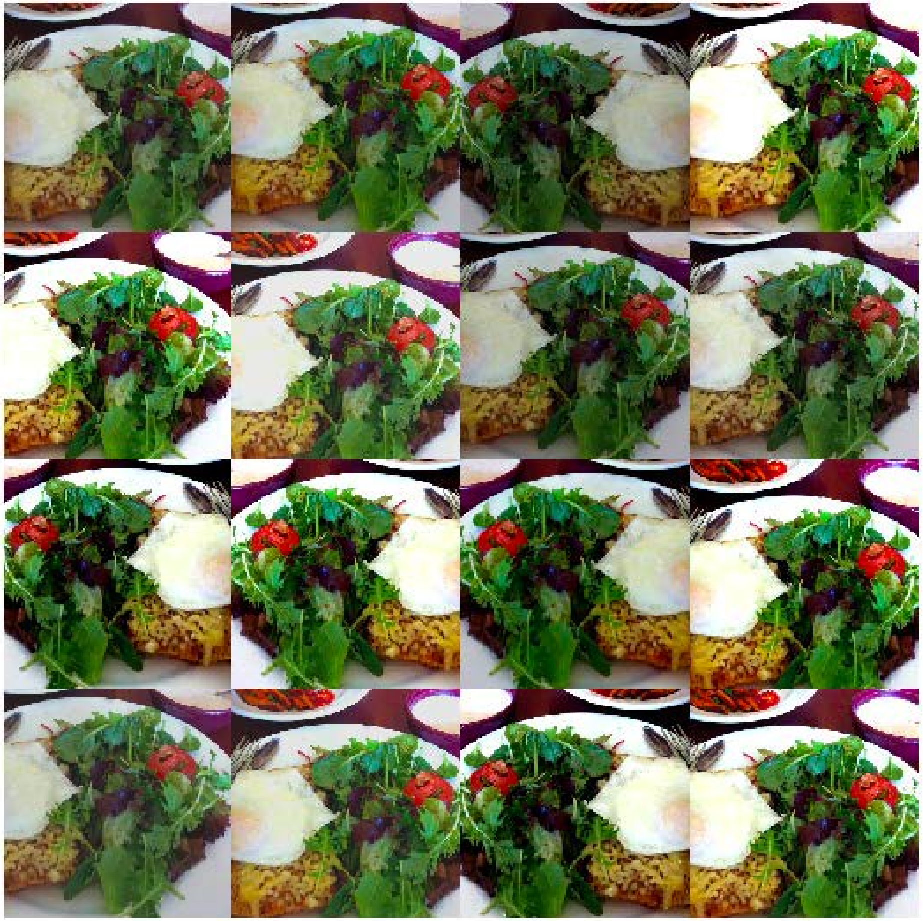


Techniques to reduce overfitting

- Reduce the number of parameters
  - Parameter sharing (convnets, recurrent neural nets)
- Early stopping
  - Indirectly controls the magnitude of the parameters
- Weight decay (aka L2 regularization)
  - Penalizes the magnitude of the parameters
- There are several other alternatives
  - Dropout, batch normalization

Against overfitting: data augmentation

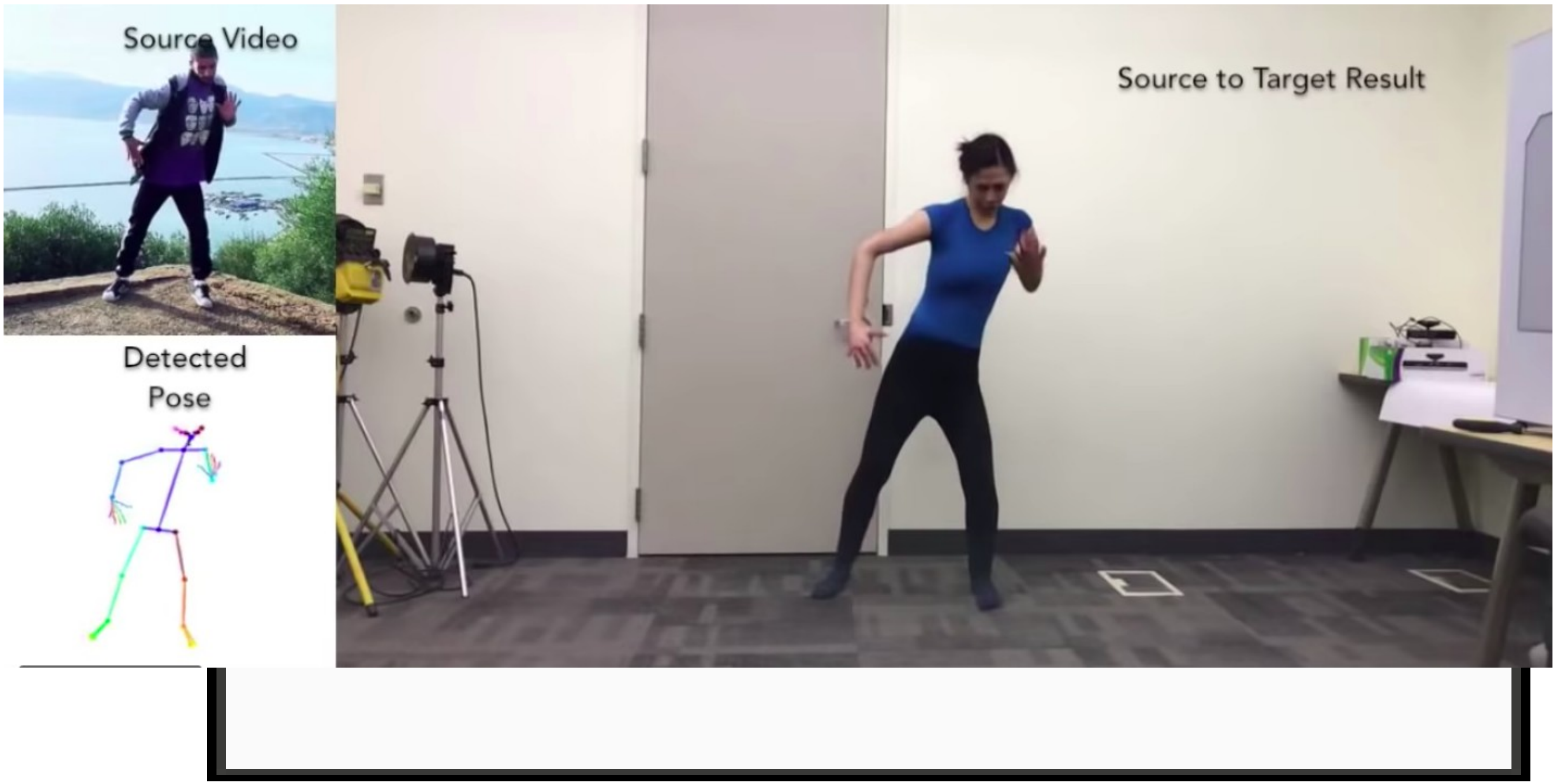
- Resizing images keeping the aspect ratio.
- Enhancing images using random distortions (color, contrast, brightness and sharpness).
- Applying random crops using the same dimension for the width and height.
- Applying random horizontal flips.



Index

- AI, Machine learning & Deep learning
- What is a Convolutional Neural Network?
  - Layers
  - Optimization
- Applications

Everybody dances now

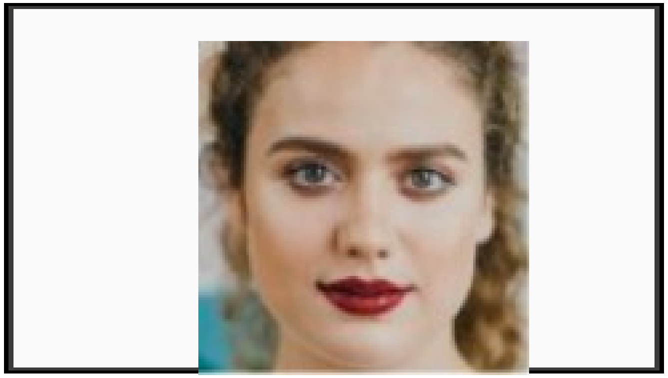


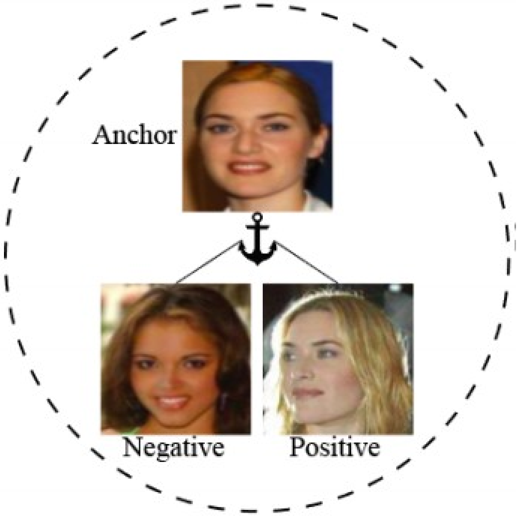
[More links]

Convolutional NN

CNN beat humans in many tasks as:

- object recognition,
- lip reading,
- high-end surveillance,
- facial recognition,
- object-based searches,
- license plate readers,
- traffic violations detection,
- breast tomosynthesis diagnosis,
- etc., etc.



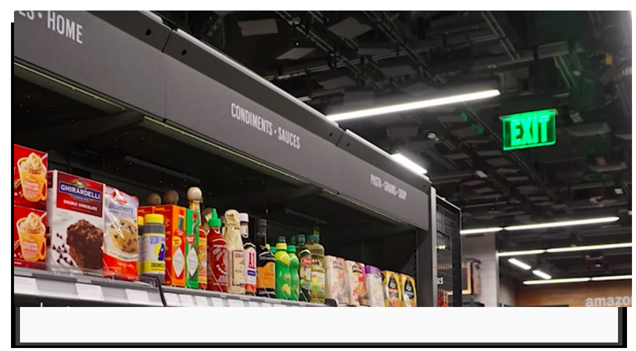


Computer Vision & Deep Learning

Social media

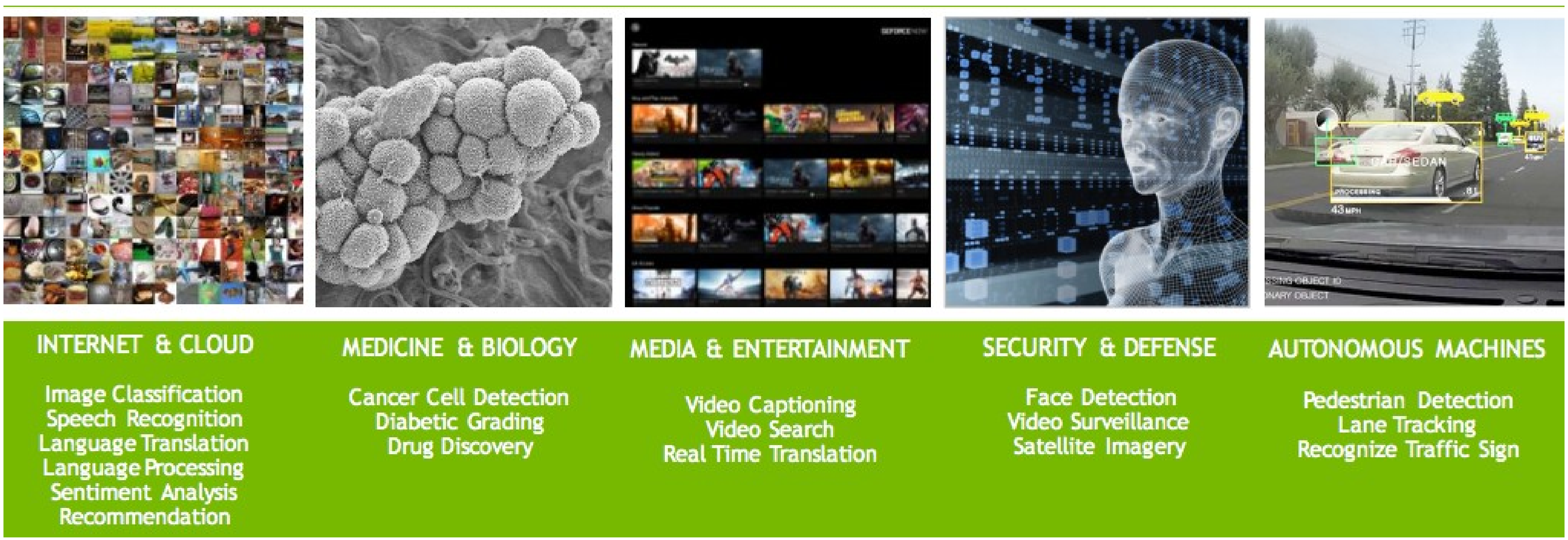
Autodriving (Tesla)

Security (Airports)



Shopping (Mango, Amazon)

Deep leearning everywhere



Deep learning - one of the 10 breakthrough technologies

MIT Technology Review

10 BREAKTHROUGH TECHNOLOGIES 2013

Introduction | The 10 Technologies | Past Years

Deep Learning

With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart.

Temporary Social Media

Messages that quickly self-destruct could enhance the privacy of online communications and make people freer to be spontaneous.

Prenatal DNA Sequencing

Reading the DNA of fetuses will be the next frontier of the genomic revolution. But do you really want to know about the genetic problems or musical aptitude of your unborn child?

Additive Manufacturing

Skeptical about 3-D printing? GE, the world's largest manufacturer, is on the verge of using the technology to make jet parts.

Baxter: The Blue-Collar Robot

Rodney Brooks's newest creation is easy to interact with, but the complex innovations behind the robot show just how hard it is to get along with people.

Memory Implants

A maverick neuroscientist believes he has deciphered the code by which the brain forms long-term memories. Next: testing a prosthetic implant for people suffering from long-term memory loss.

Smart Watches

The designers of the Pebble watch realized that a mobile phone is more useful if you don't have to take it out of your pocket.

Ultra-Efficient Solar Power

Doubling the efficiency of a solar cell would completely change the economics of renewable energy. Nanotechnology just might make it possible.

Big Data from Cheap Phones

Collecting and analyzing information from simple cell phones can provide surprising insights into how people move about and behave – and even help us understand the spread of diseases.

Supergrids

A new high-power circuit breaker could finally make highly efficient DC power grids practical.

Deep leearning everywhere

"Preparing for the Future of Artificial Intelligence." ★1

This 58-page report outlines a number of important topics related to artificial intelligence.

http://readwrite.com/2016/10/16/white-house-offers-artificial-intelligence-plan-cl1/

AI & ML: from R+D lab to hospital

mobihealthnews

TOPICS MAIN MENU

APAC EMEA Global Edition

FDA permits marketing of AI software that autonomously detects diabetic retinopathy

By **Dave Muoio** | April 12, 2018 | 11:10 am

SHARE Share 733

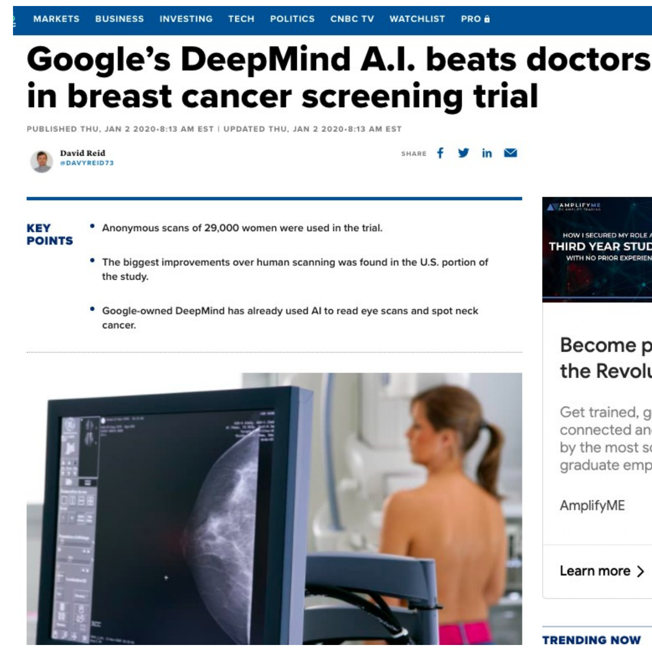
The FDA has granted diagnostics company IDx's De Novo request to market its AI-based software system for the autonomous detection of diabetic retinopathy in adults who have diabetes, called IDx-DR.

This decision represents the first AI-based diagnostic system authorized by the FDA for commercialization in the US that can provide a screening decision without the need for clinician interpretation, according to the agency. The news comes just months after the **FDA's De Novo approval of Viz.AI**, another AI software tool that analyzes stroke indicators and highlights CT images that could require additional clinical attention.

"Early detection of retinopathy is an important part of managing care for the millions of people with diabetes, yet many patients with diabetes are not adequately screened for diabetic retinopathy since about 50 percent of them do not see their eye doctor on a yearly basis," Dr. Malvina Eydelman, director of the Division of Ophthalmic and Ear, Nose, and Throat Devices at the FDA's Center for Devices and Radiological Health, said in a statement. "Today's decision permits the marketing of a novel artificial

(China Daily) 14:15, July 02, 2018 Follow on Apple News

Radiologist Zhang Junhai from Shanghai Huashan Hospital reads a medical image display during a competition with BioMind, an artificial intelligence system, in Beijing



MARKETS BUSINESS INVESTING TECH POLITICS CNBC TV WATCHLIST PRO

Google's DeepMind A.I. beats doctors in breast cancer screening trial

PUBLISHED THU, JAN 2 2020·8:13 AM EST | UPDATED THU, JAN 2 2020·8:13 AM EST

David Reid
@DAVYREID73

SHARE

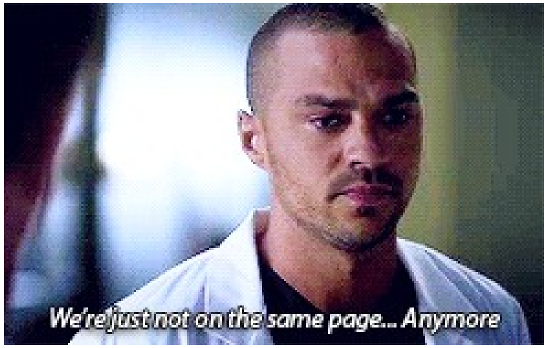
KEY POINTS

- Anonymous scans of 29,000 women were used in the trial.
- The biggest improvements over human scanning was found in the U.S. portion of the study.
- Google-owned DeepMind has already used AI to read eye scans and spot neck cancer.

TRENDING NOW

Concerns about DL

- Different development approach.



- No clear view on how insight is generated.

- A system is only as good as the data it learns from.

Read more at https://indatalabs.com/blog/data-science/problems-of-artificial-intelligence-implementation#hUdVxpB3pFsMqgdF.99

Conclusions

Deep learning – a technology that came to stay

A new technological trend that is affecting directly our environment

4 kind of layeres (convolutional, RELU, maxpooling and fully connected) form any neural network.

The loss function controls the performance, small loss function means high performing score function

Overfitting is one of the main problems due to the millions of parameters.

Stochastic gradient descent and similar approaches are used to optimize the loss function.

Different CNN models can be found. No optimal one exists.

CNN applicable to different Computer Vision problems

What are the next challenges?

1. Uncertainty modeling
   - Why are CNNs so efficient? Optimal?
   - Inference: In many serious applications, we need error bar
2. Interpretability and explainability: causality models
3. Unsupervised and Few-shot learning
4. Dataset models
   1. Data are not free,
   2. GDPR doesn't hellp always
5. GPUs-hungry research
   1. Energy-overconsuming
6. Technology is still very young
7. Talent deficit