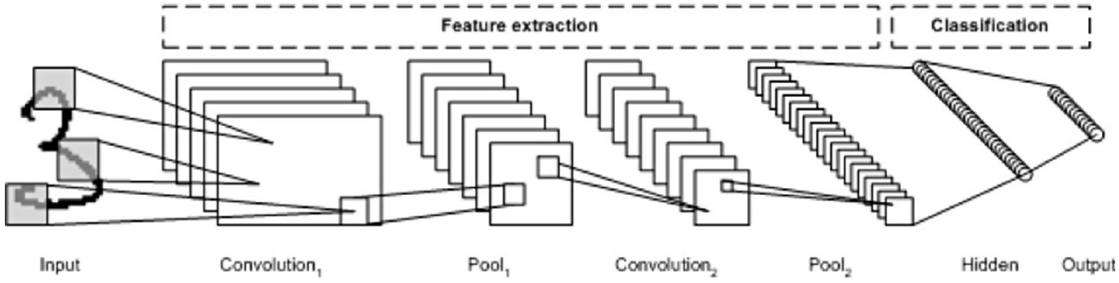
Computational vision:

Convolutional Neural Networks

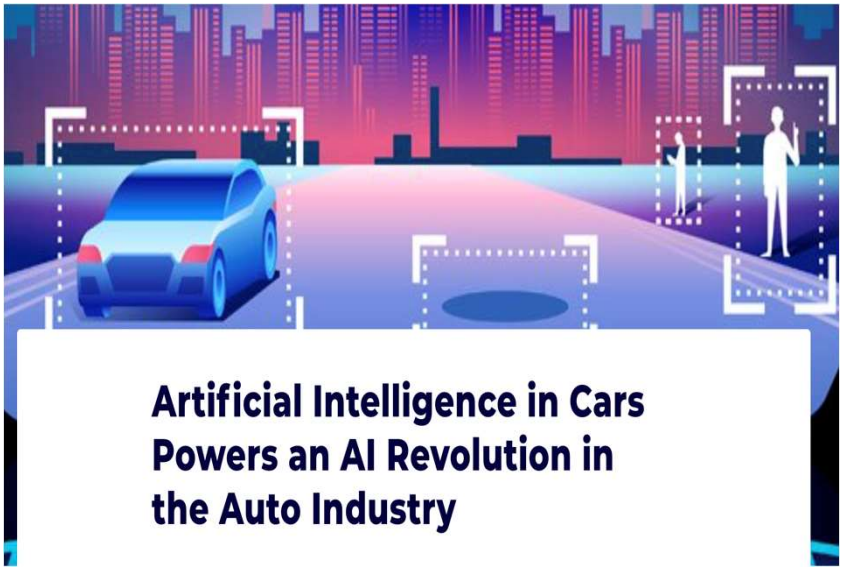
Class 6: Computational Vision

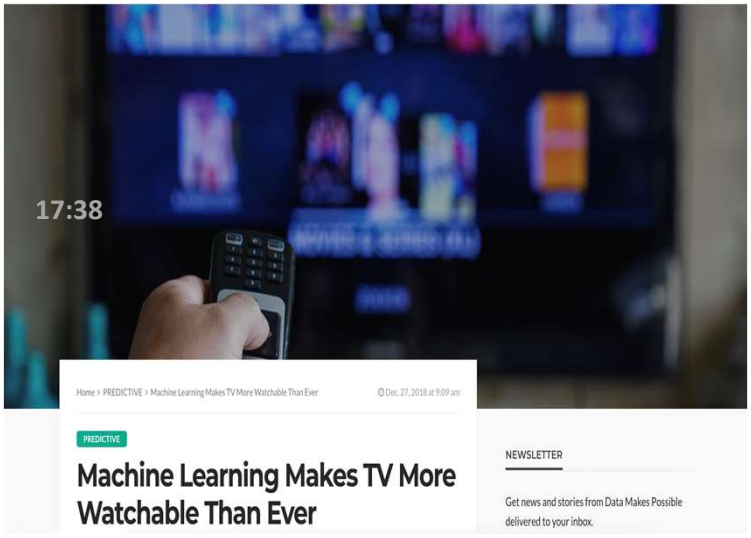


Index

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

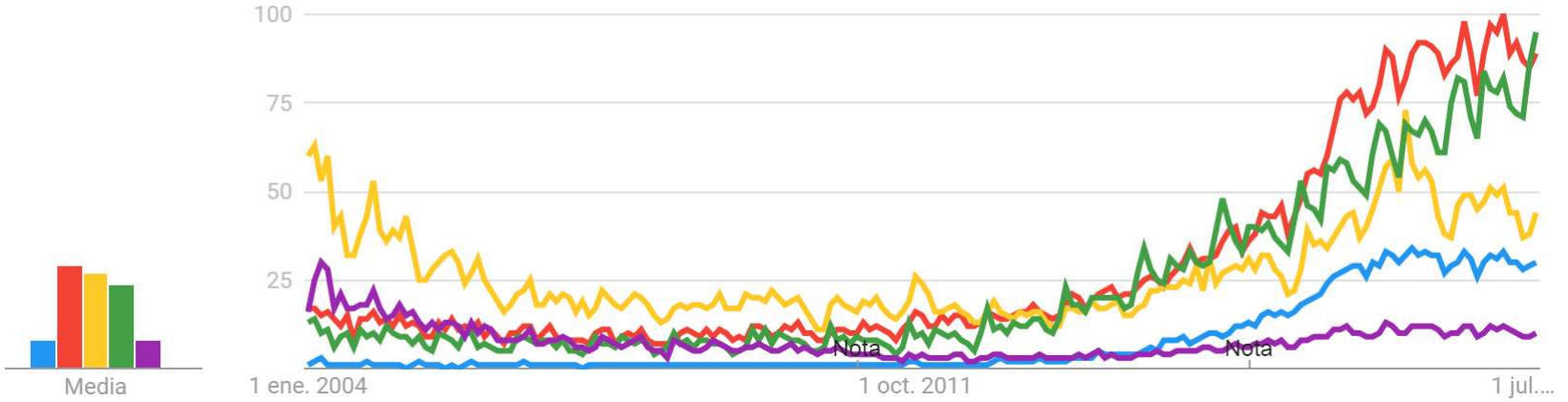
Deep learning everywhere



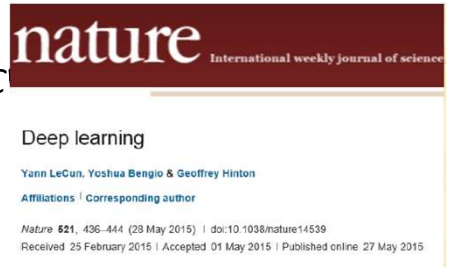


Artificial Intelligence

- Artificial Intelligence
- Machine Learning
- Deep learning
- Neural networks
- Data science

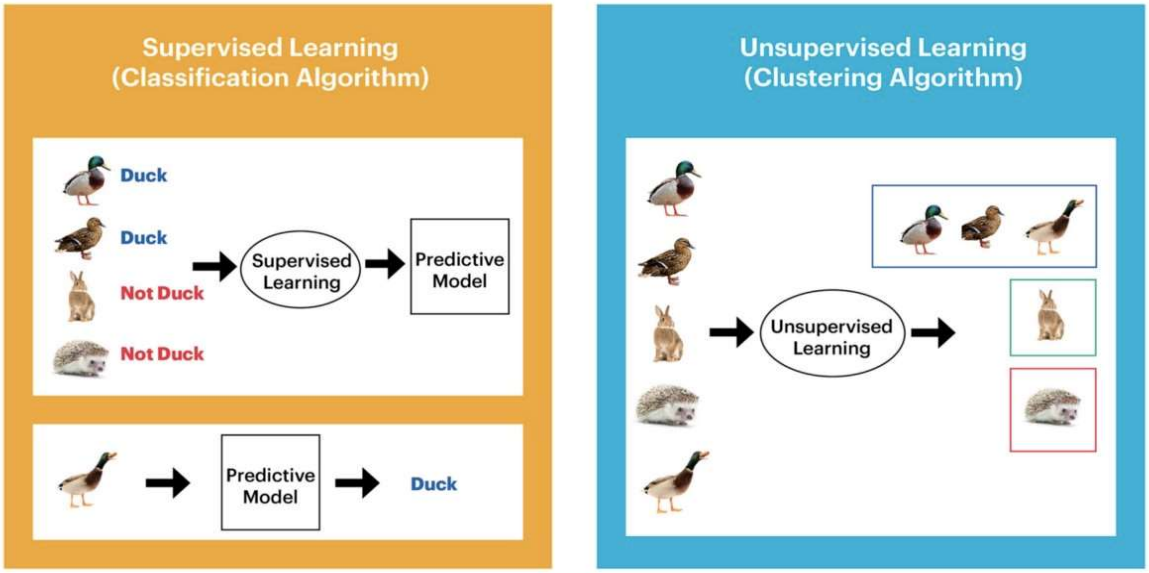


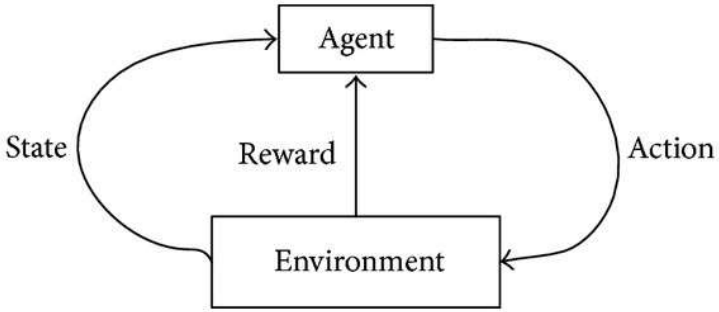
Google Scholar reveals its most influential papers



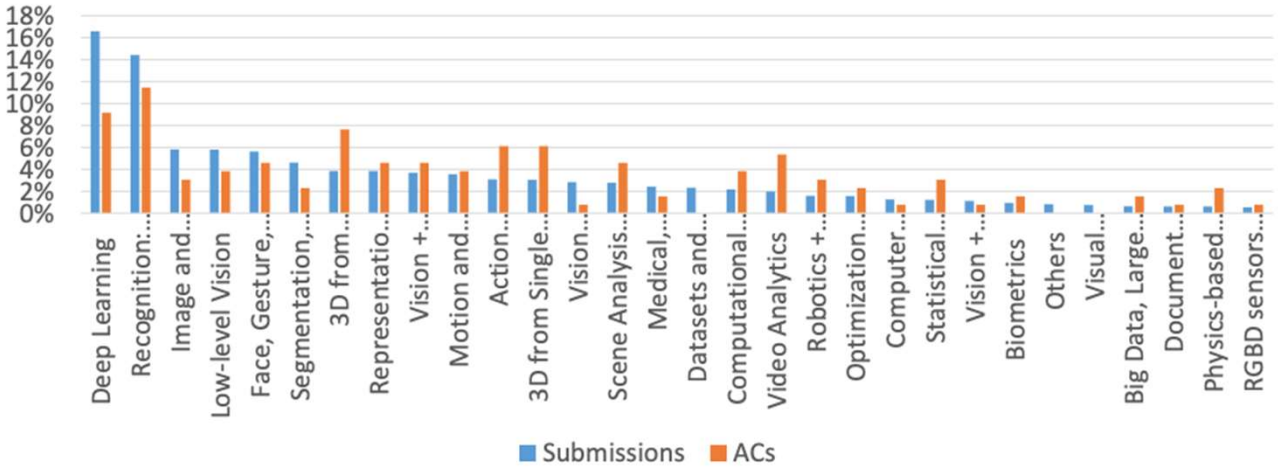
1. **"Deep Residual Learning for Image Recognition" (2016)** *Proceedings of the IEEE/C Computer Vision and Pattern Recognition* 25,256 citations
2. **"Deep learning" (2015)** *Nature* 16,750 citations
3. **"Going Deeper with Convolutions" (2015)** *Proceedings of the IEEE/CVF Conference on Computer 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

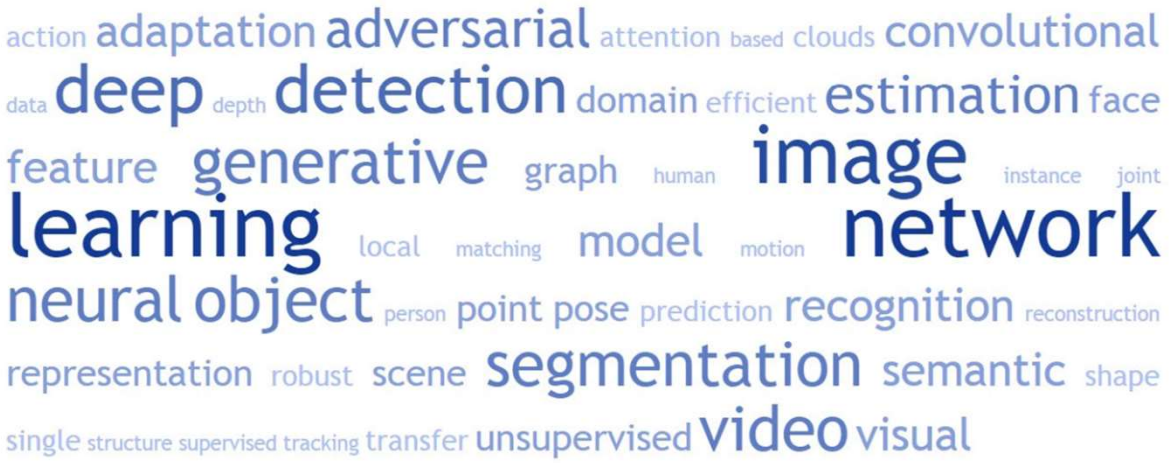
Supervised vs. Unsupervised vs Reinforcement Learning





CVPR'2019


18%
16%
14%
12%
10%
8%
6%
4%
2%
0%
Deep Learning
Recognition:...
Image and...
Low-level Vision
Face, Gesture,...
Segmentation,...
3D from...
Representatio...
Vision +...
Motion and...
Action...
3D from Single...
Vision...
Scene Analysis...
Medical,...
Datasets and...
Computational...
Video Analytics
Robotics +...
Optimization...
Computer...
Statistical...
Vision +...
Biometrics
Others
Visual...
Big Data, Large...
Document...
Physics-based...
RGBD sensors...
Submissions
ACs


action adaptation adversarial attention based clouds convolutional
data deep depth detection domain efficient estimation face
feature generative graph human image instance joint
learning local matching model motion network
neural object person point pose prediction recognition reconstruction
representation robust scene segmentation semantic shape
single structure supervised tracking transfer unsupervised video visual

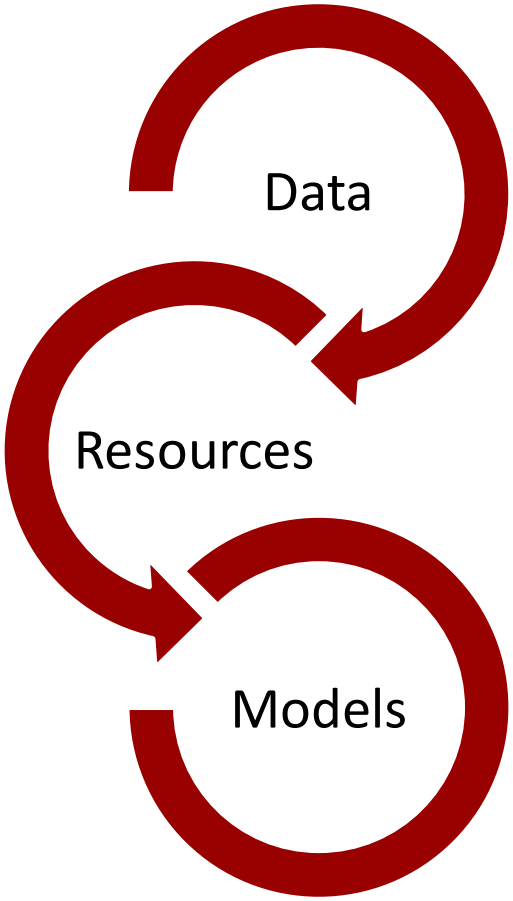
Deep learning – where does it come from?

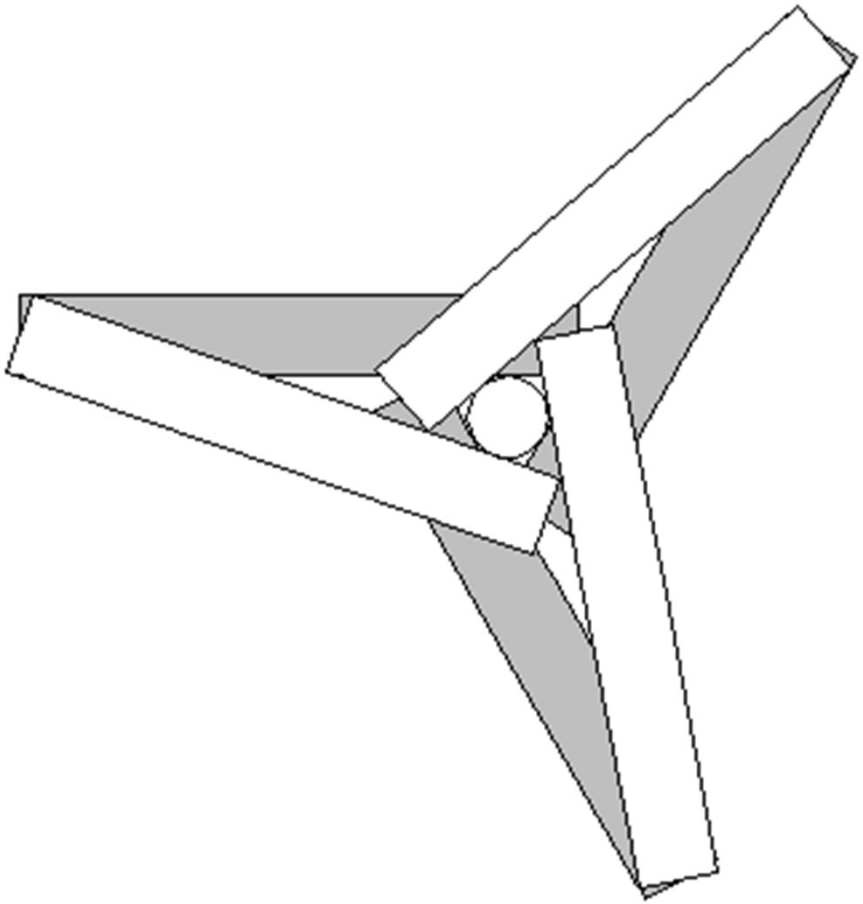
- **1943:** neurophysiologist Warren McCulloch and mathematician Walter Pitts – a **neuronal model** using an electrical circuit
- **1950**: Alan Turing created the world-famous **Turing Test**: a computer to pass, has to be able to convince a human that it is a human and not a computer.
- **1952**: 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





Data

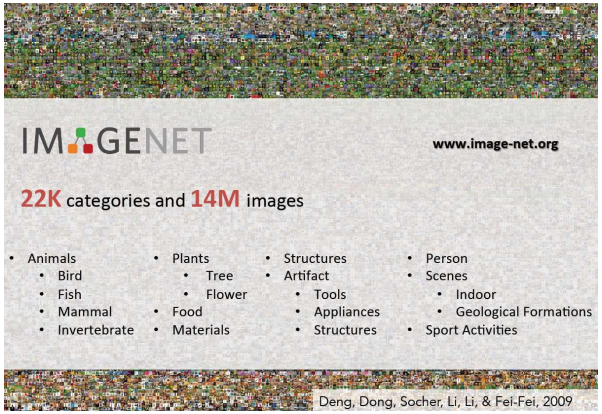
90% of all digital data were generated last 2 years.

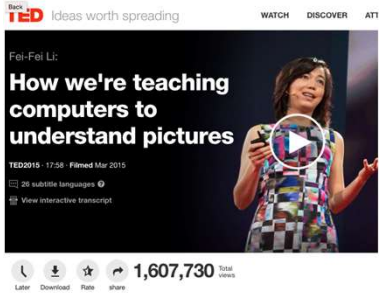
Every minute of the day:

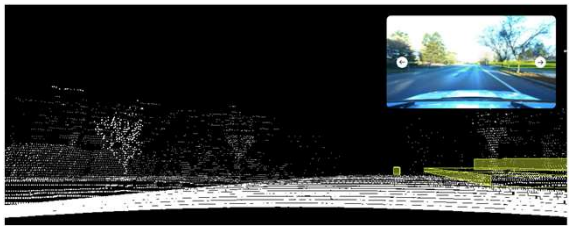
- 4M YouTube videos watched
- 456K tweets on Twitter
- 46K potos posted in Instagram
- 16M text messages sent
- 103M spam emails sent

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



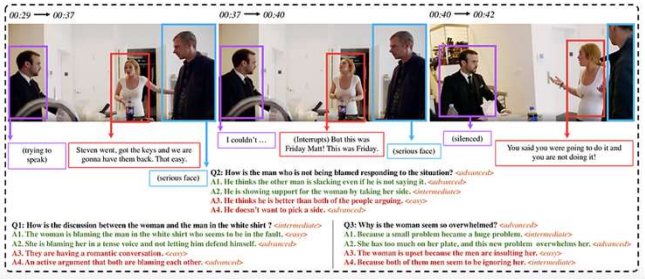




Lyft Level 5

LVIS Challenge:
2.2M masks, 16K images

Places2: 10M images



SocialIQ

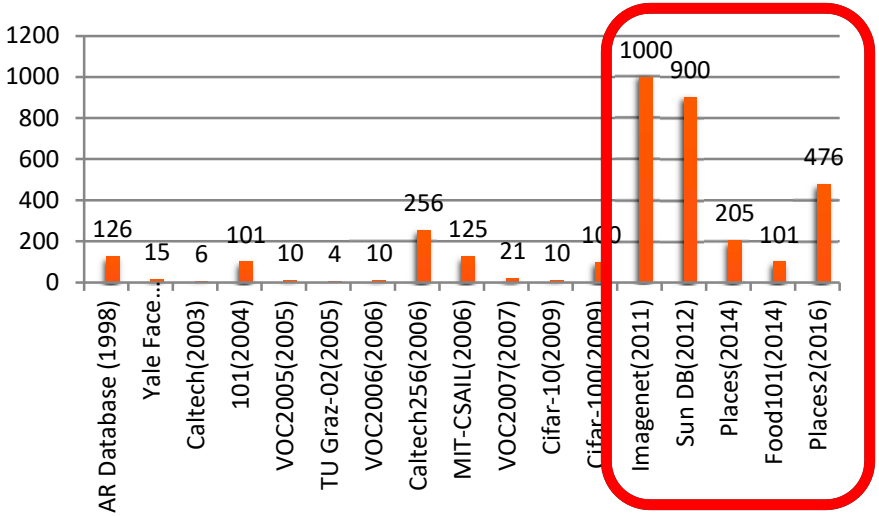
TACO: Waste in the wild

FastMRI

https://www.datasetlist.com

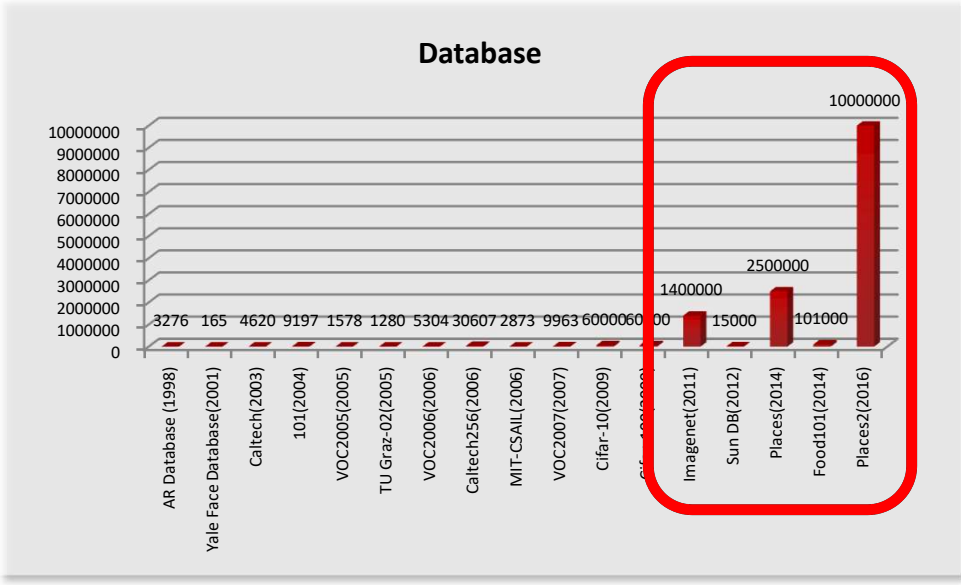
Image databases evolution

Number of objects/Database

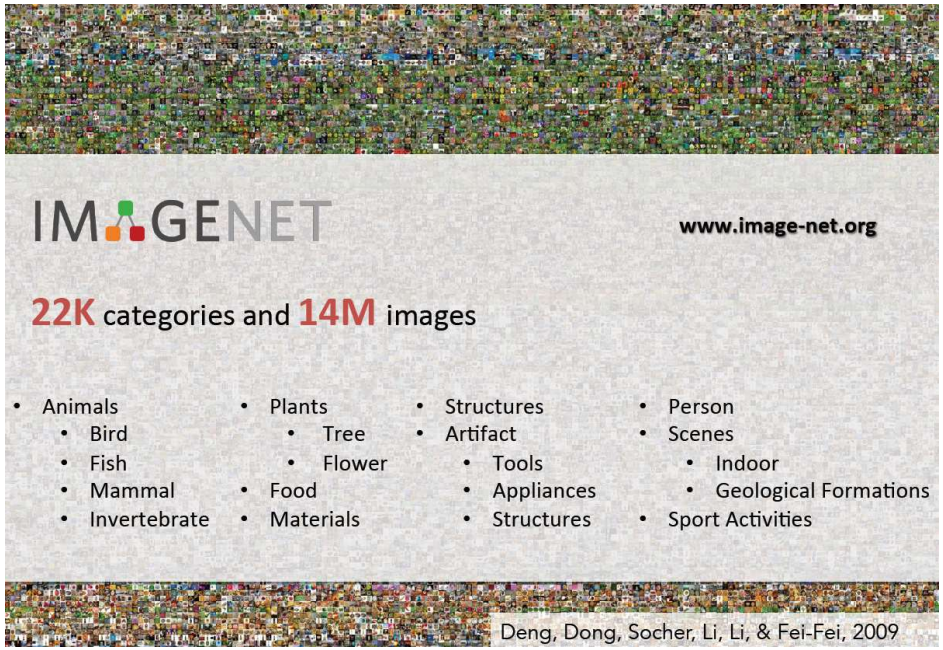


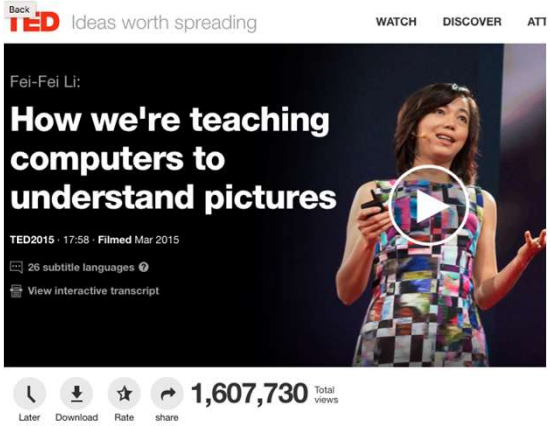
ImageNet & Deep learning

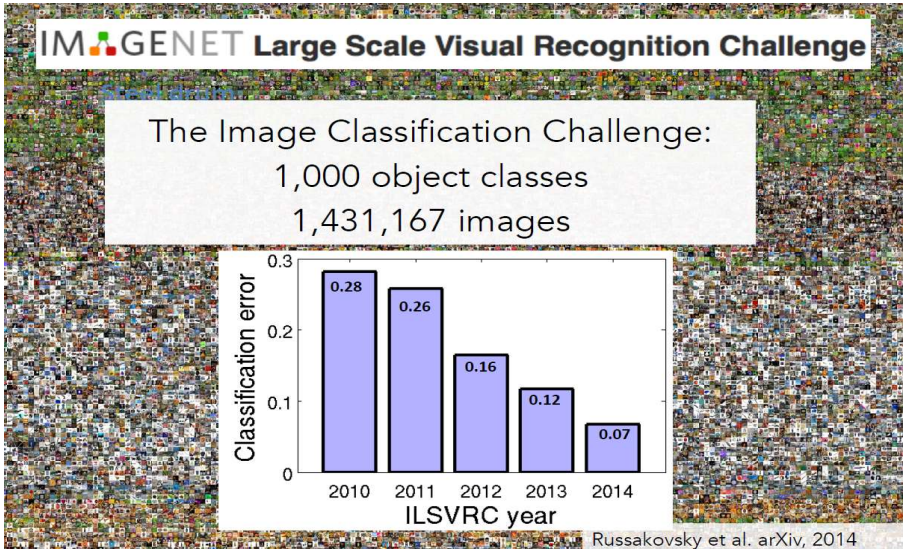
Number of images/Database



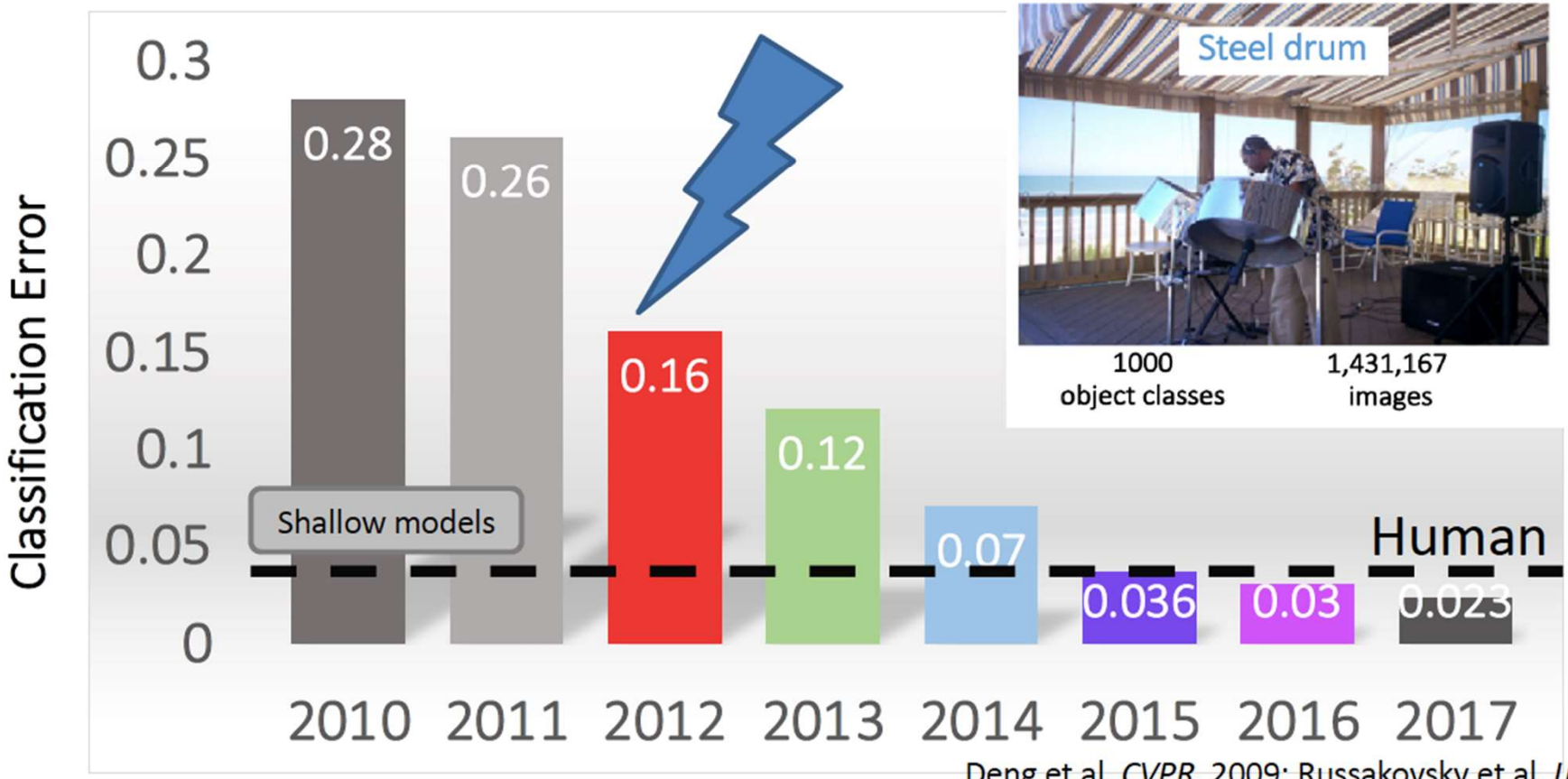
Imagenet


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
Deng, Dong, Socher, Li, Li, & Fei-Fei, 2009


Ideas worth spreading
WATCH
DISCOVER
Fei-Fei Li:
How we're teaching computers to understand pictures
TED2015 · 17:58 · Filmed Mar 2015
26 subtitle languages
View interactive transcript
Later
Download
Rate
share
1,607,730 Total views


IMAGENET Large Scale Visual Recognition Challenge
The Image Classification Challenge:
1,000 object classes
1,431,167 images
Classification error
0.3
0.2
0.1
0
0.28
0.26
0.16
0.12
0.07
2010
2011
2012
2013
2014
ILSVRC year
Russakovsky et al. arXiv, 2014

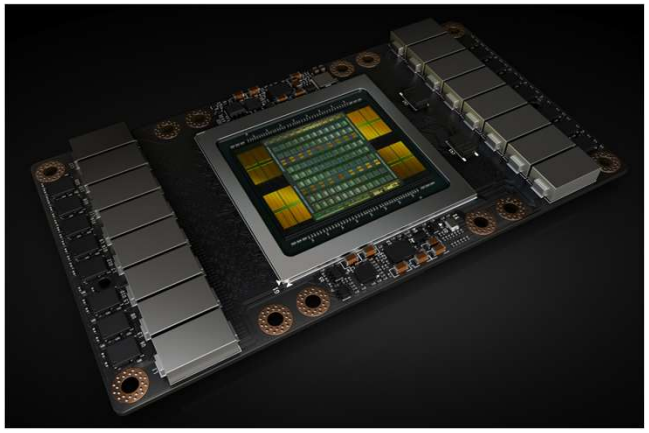
IMAGENET Classification Task



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)



$$D = \begin{pmatrix} A_{0,0} & A_{0,1} & A_{0,2} & A_{0,3} \\ A_{1,0} & A_{1,1} & A_{1,2} & A_{1,3} \\ A_{2,0} & A_{2,1} & A_{2,2} & A_{2,3} \\ A_{3,0} & A_{3,1} & A_{3,2} & A_{3,3} \end{pmatrix} \begin{pmatrix} B_{0,0} & B_{0,1} & B_{0,2} & B_{0,3} \\ B_{1,0} & B_{1,1} & B_{1,2} & B_{1,3} \\ B_{2,0} & B_{2,1} & B_{2,2} & B_{2,3} \\ B_{3,0} & B_{3,1} & B_{3,2} & B_{3,3} \end{pmatrix} + \begin{pmatrix} C_{0,0} & C_{0,1} & C_{0,2} & C_{0,3} \\ C_{1,0} & C_{1,1} & C_{1,2} & C_{1,3} \\ C_{2,0} & C_{2,1} & C_{2,2} & C_{2,3} \\ C_{3,0} & C_{3,1} & C_{3,2} & C_{3,3} \end{pmatrix}$$

FP16 or FP32 FP16 FP16 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

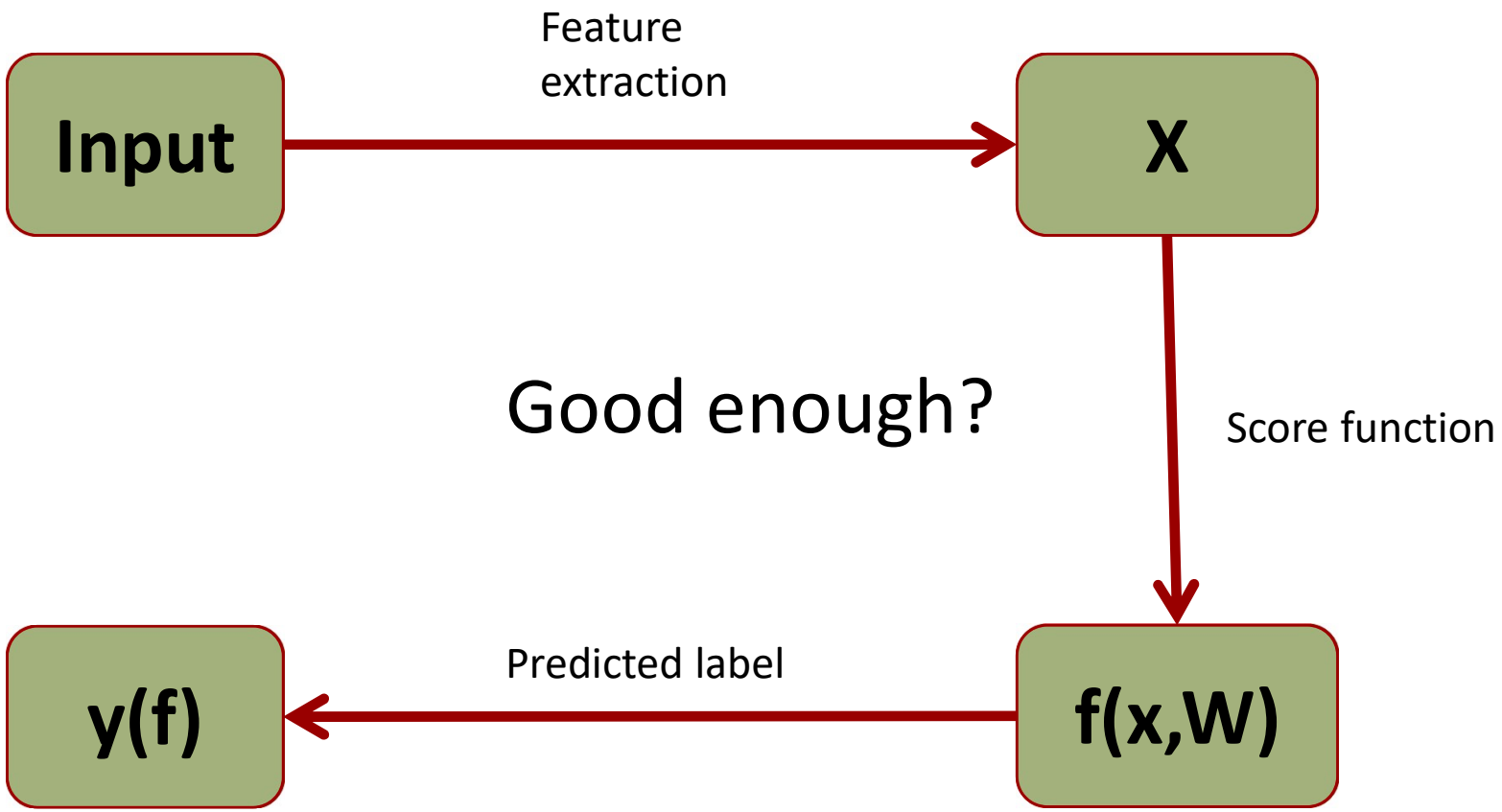
- AI, Machine learning & Deep learning
- What is a Convolutional Neural Network?
  - Layers
  - Optimization
- Applications

Why Deep learning?

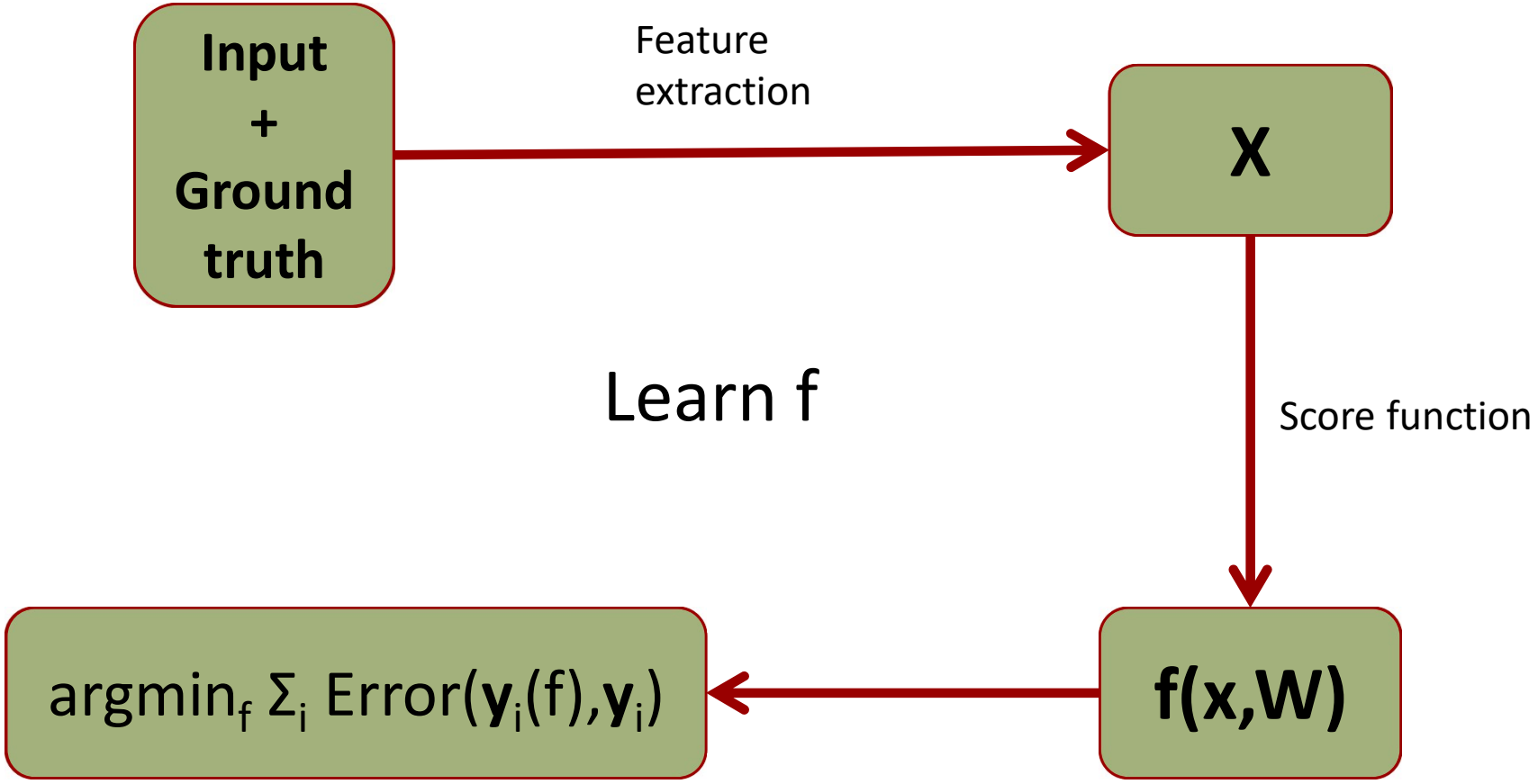
Classical way of solving Computer Vision problems

How Computer Vision problems are solved by Deep Learning

The learning pipeline



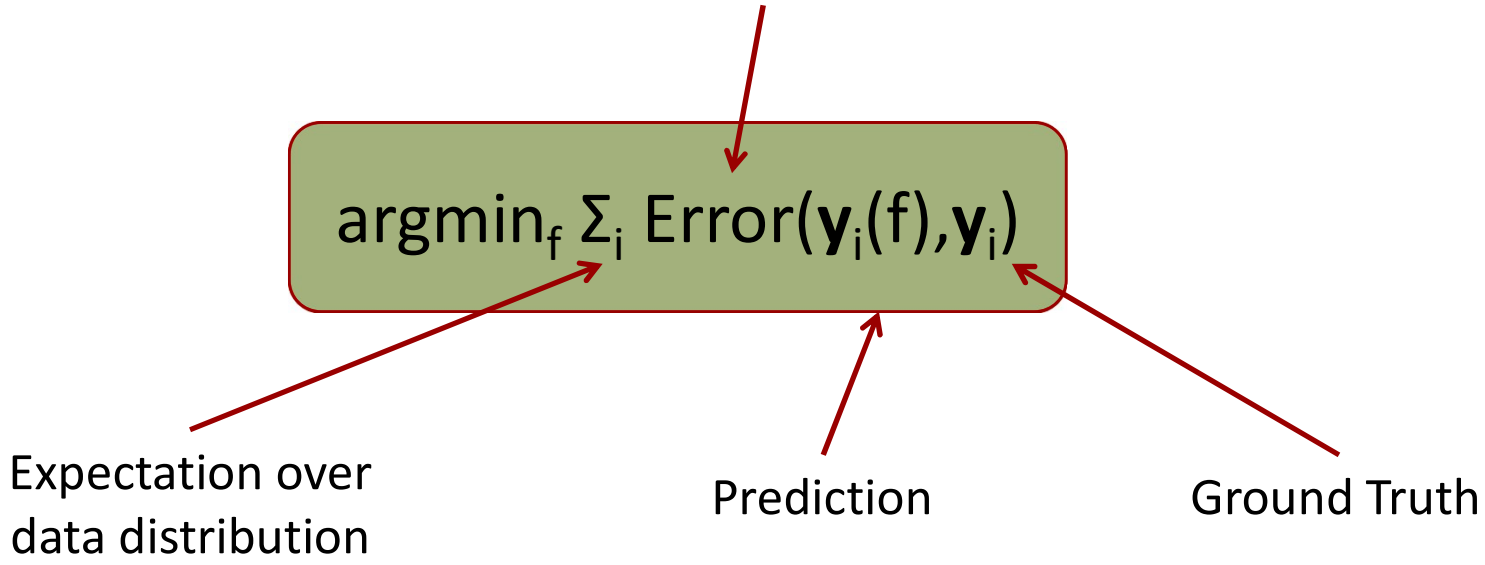
The traning process



The learning process

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

Measure of prediction quality (error, loss)

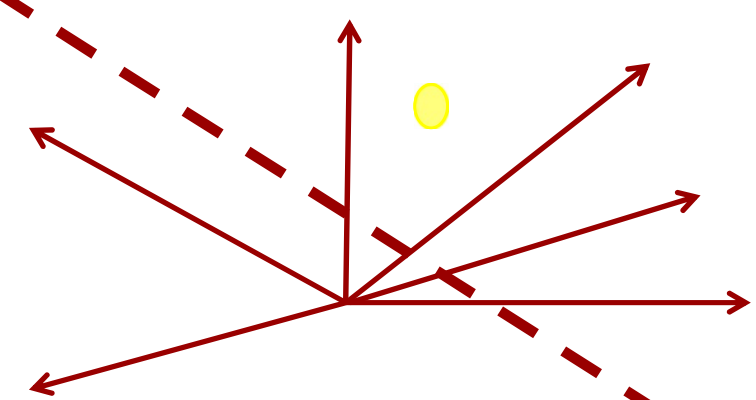


Loss function the negative conditional log-likelihood, with the interpretation that $f_i(X)$ estimates $P(Y=i|X)$:

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

Linear classification

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



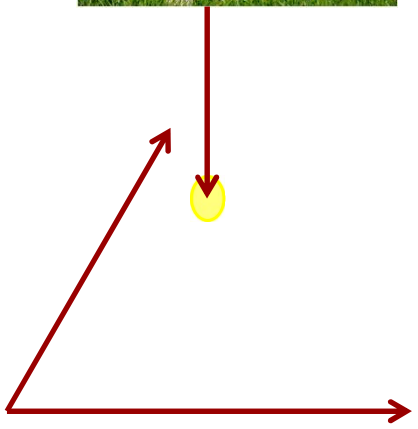
$R^{32x32x3}$

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

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

Linear classification

How to project data in the feature space:



3x1 3072x1

$$f(x) = W\,x + b$$

3x3072 3x1

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

The matrix W is (3x3072).

The bias vector b is 3-dimensional.

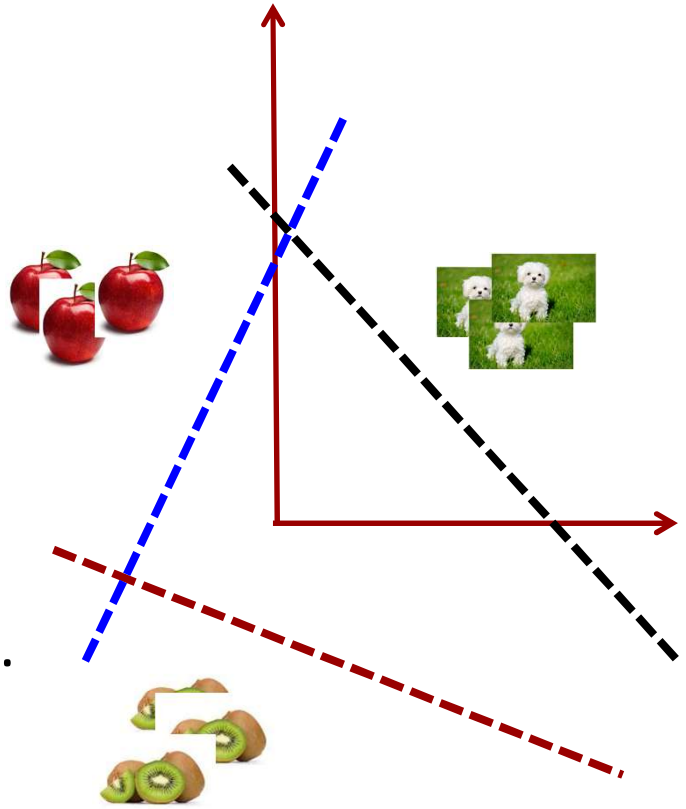
Linear classification

How to project data in the feature space:

3x1 3072x1

$$f(x) = W\,x + b$$

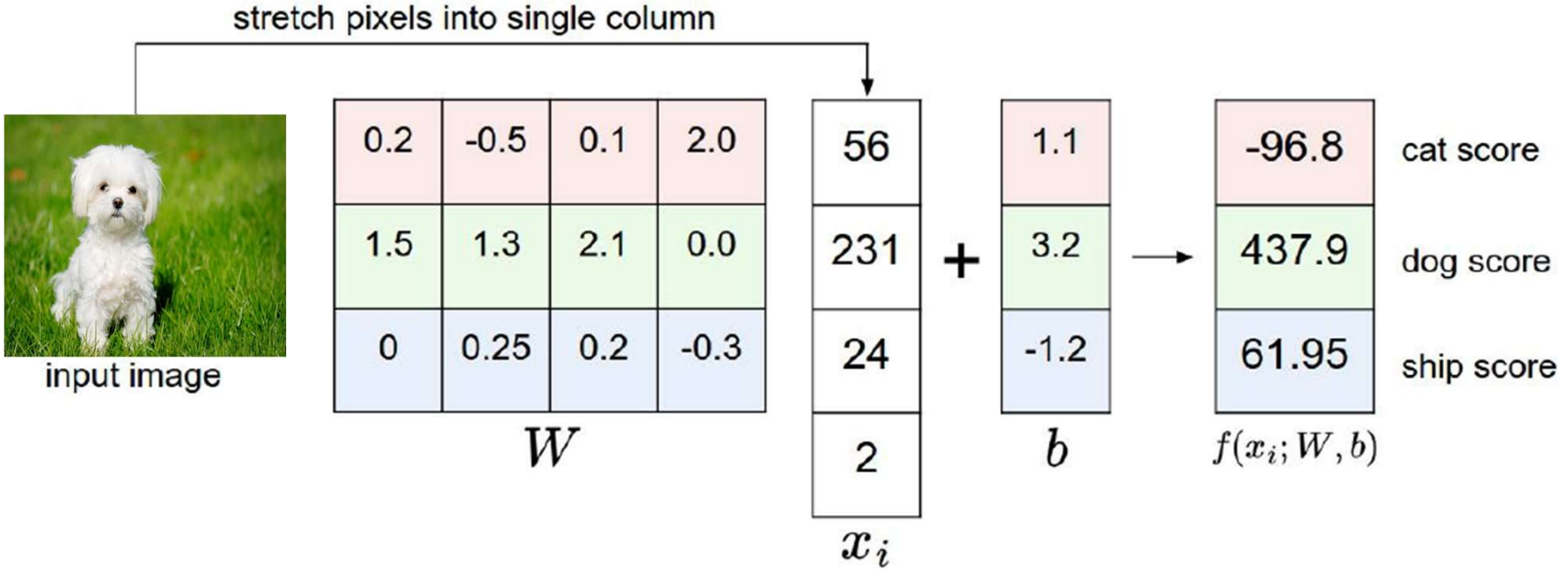
3x3072 3x1



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)

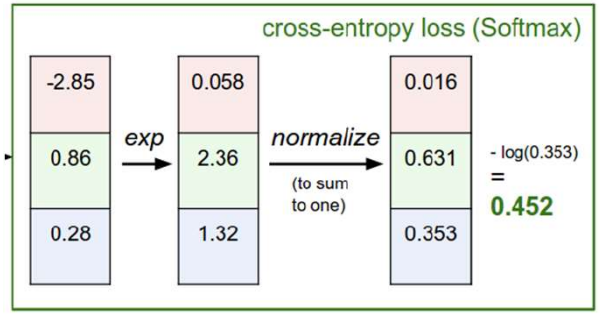


Loss function and optimisation

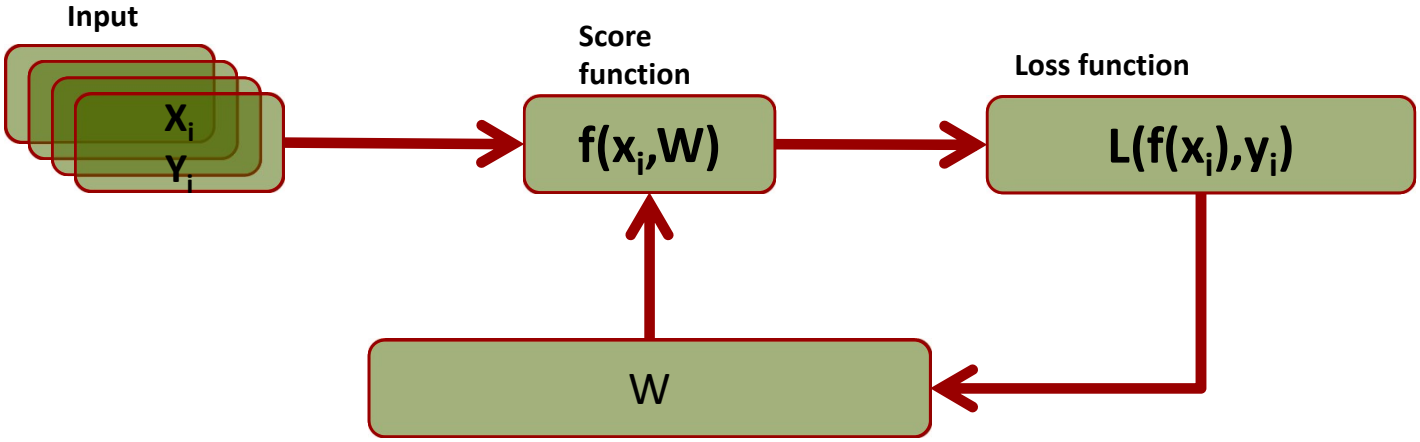
Question: if you were to assign a single number to how unhappy you are with these scores, what would you do?

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

softmax function

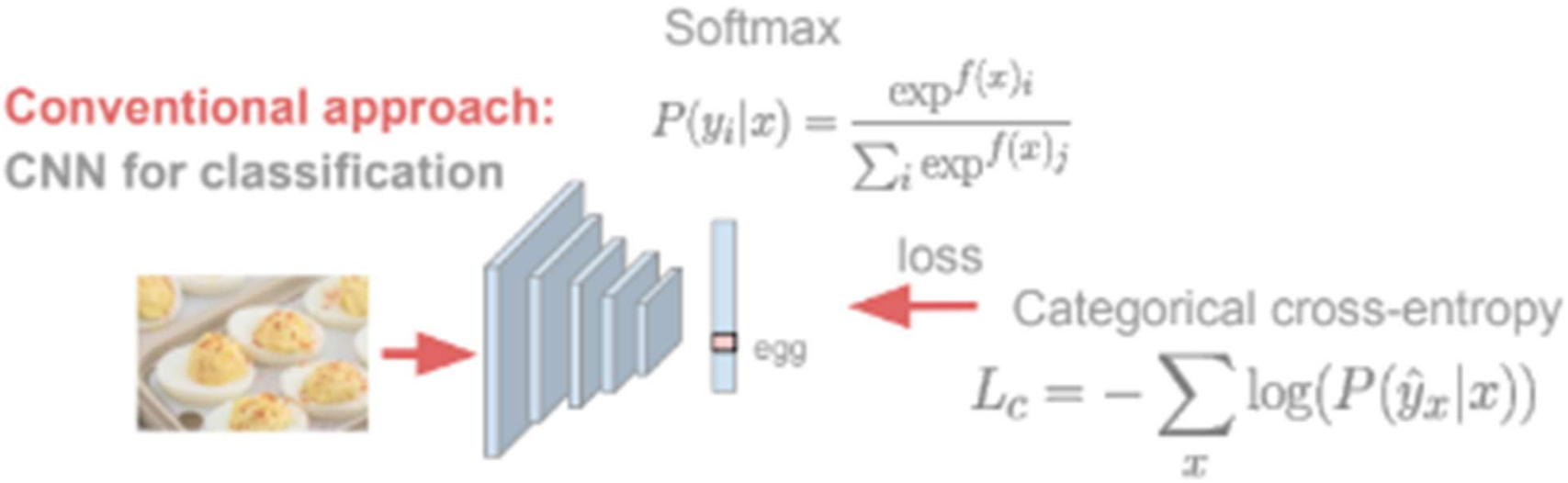


Question : Given the score and the loss function, how to find the parameters W?



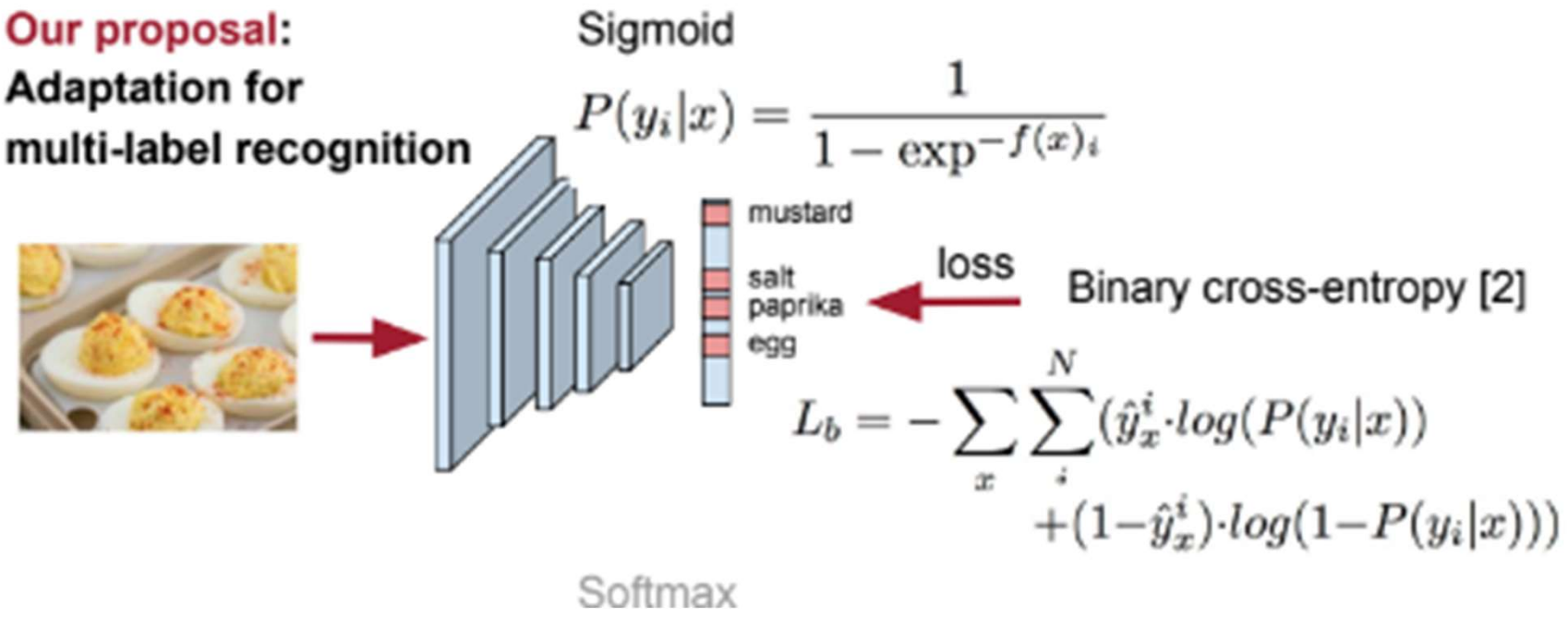
Single-label classification

- 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:

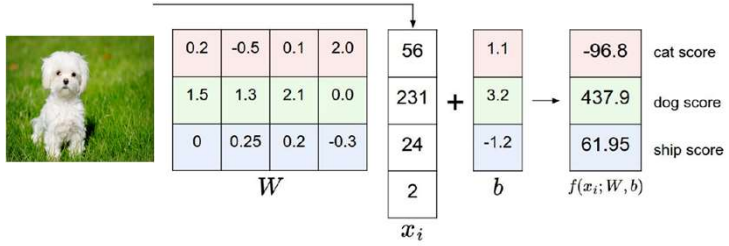


Index

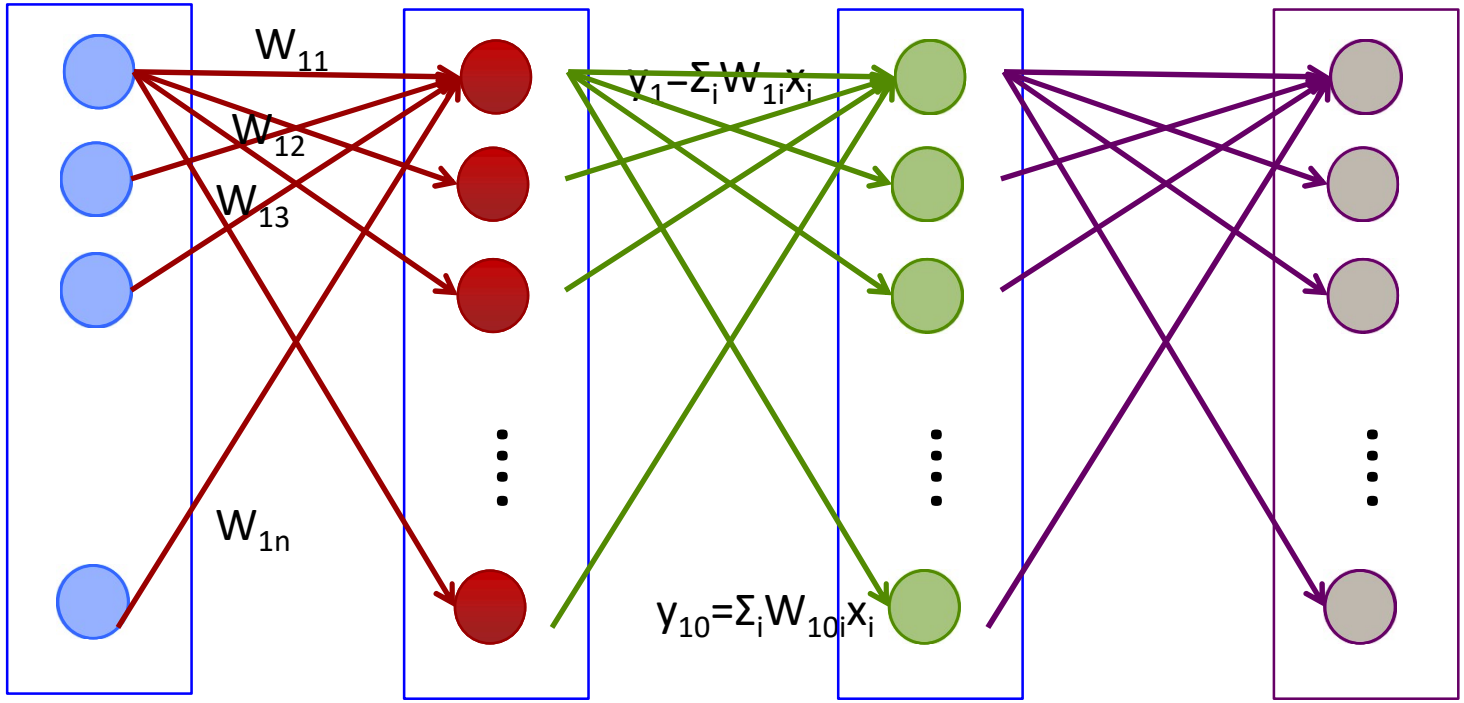
- AI, Machine learning & Deep learning
- What is a Convolutional Neural Network?
  - Layers
  - Optimization
- Applications

How is a CNN doing deep learning?

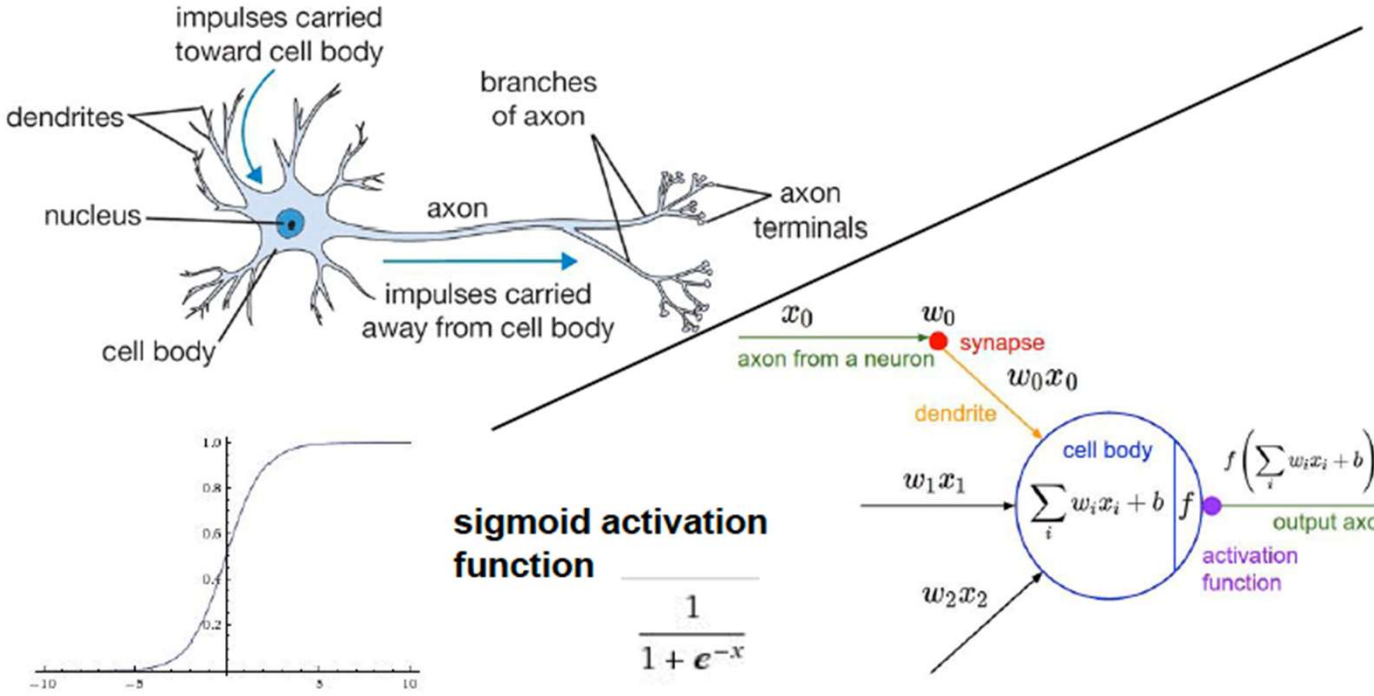
$y=Wx$ $y_1=\Sigma_i W_{1i}x_i$ ... $y=W(Wx)$ $y=W(W(Wx))$



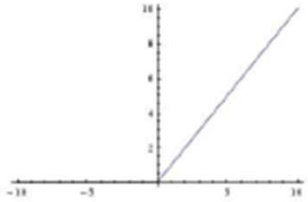
Fully connected layers



Why a CNN is a neural network?



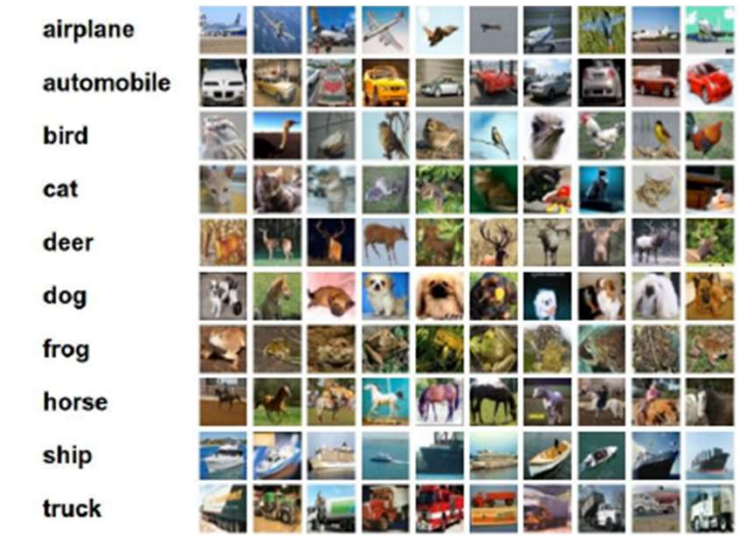
ReLU max(0,x)



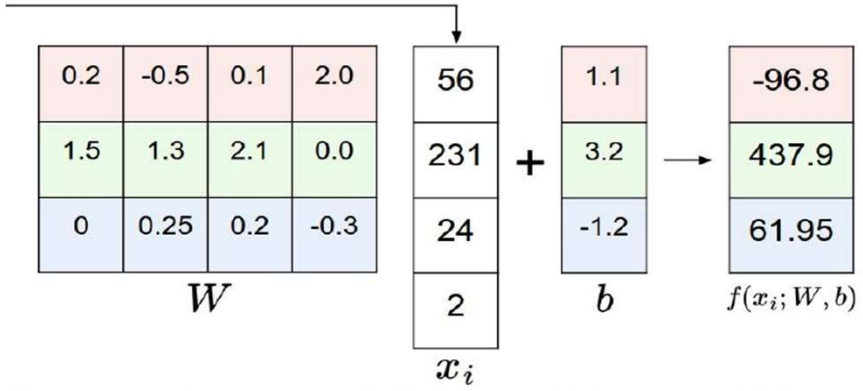
Modern CNNs – 10M neurons
Human CNNs – 5B of neurons.

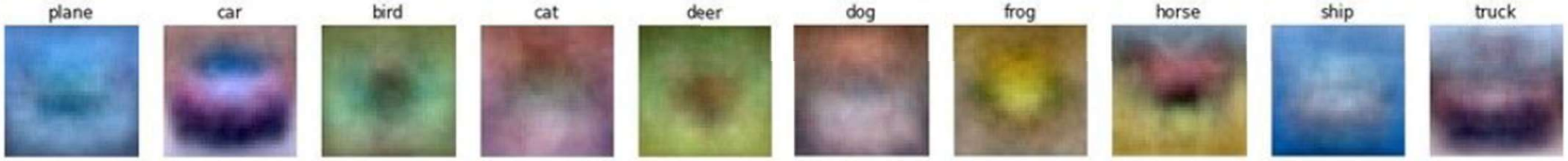
From: Fei-Fei Li & Andrej Karpathy & Justin Johnson

Why is it convolutional?



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



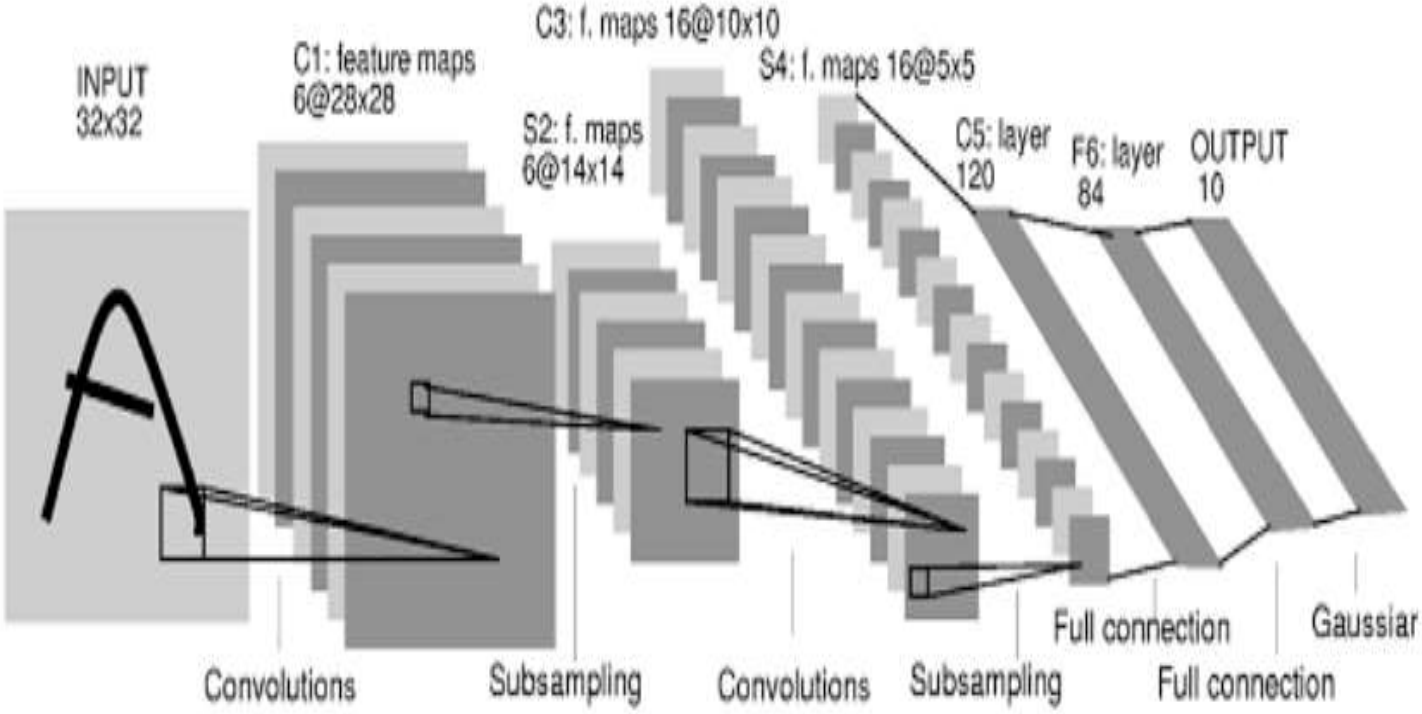


Adapted from: Fei-Fei Li & Andrej Karpathy & Justin Johnson

What is new in the Convolutional Neural Network?

1998

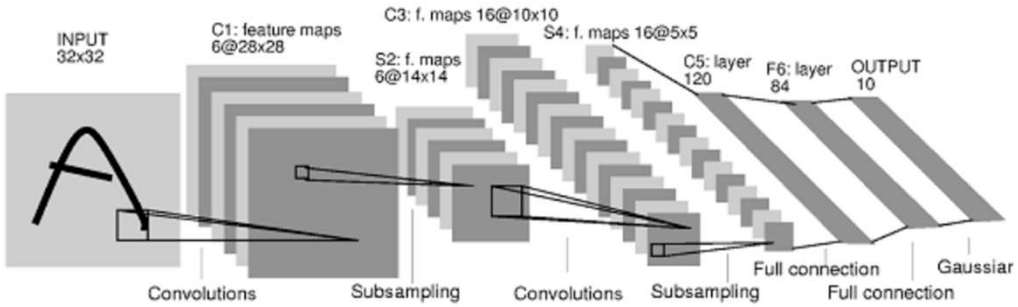
LeCun et al.



CNN evolution

1998

LeCun et al.



# of transistors

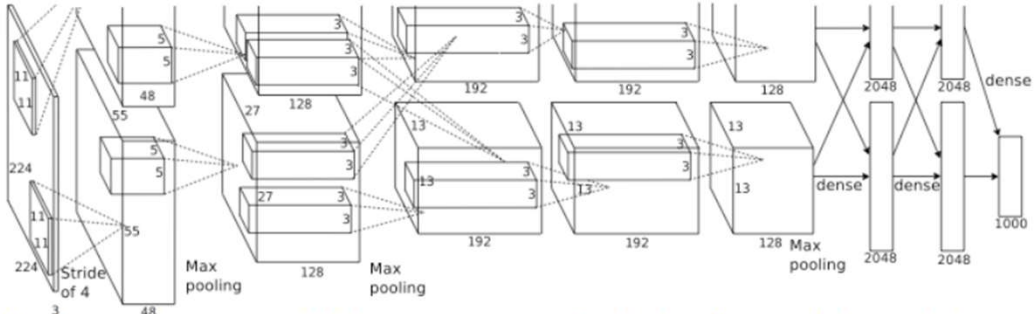
10^6

# of pixels used in training

10^7 NIST

2012

Krizhevsky et al.



# of transistors

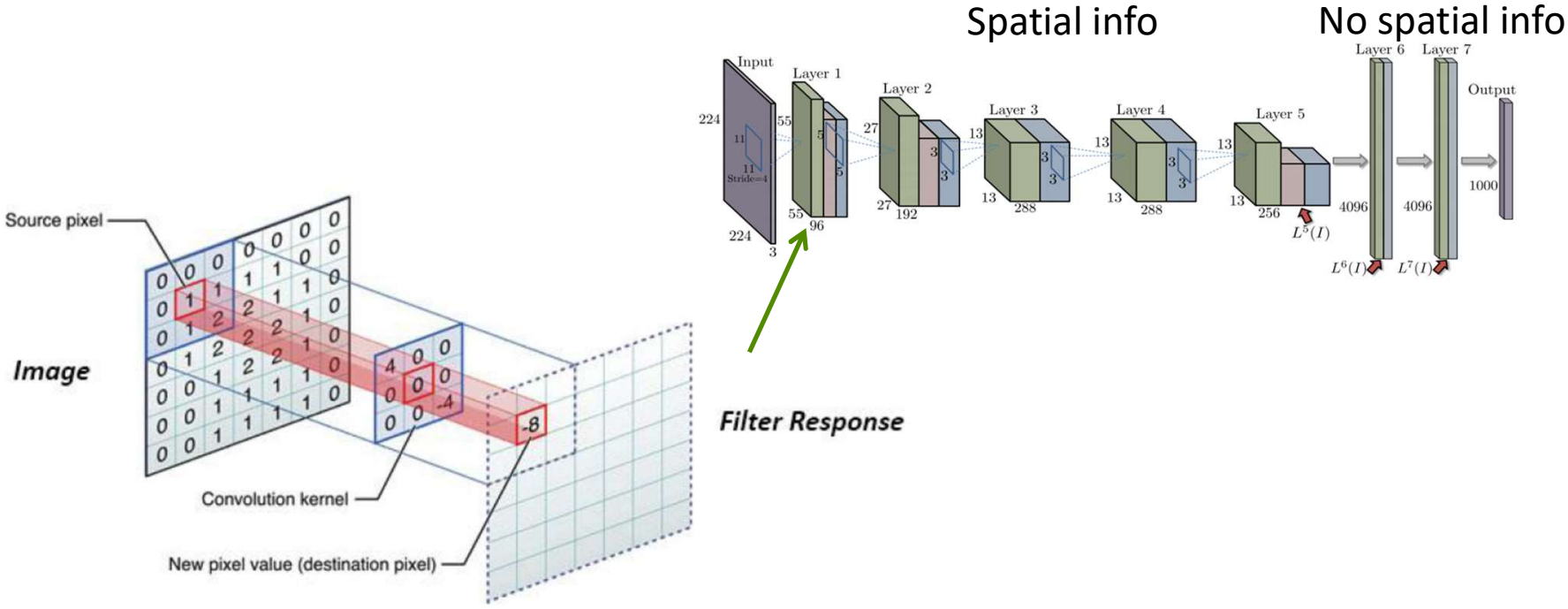
10^9

GPUs

# of pixels used in training

10^{14} IMAGENET

Convolutional and Max-pooling layer



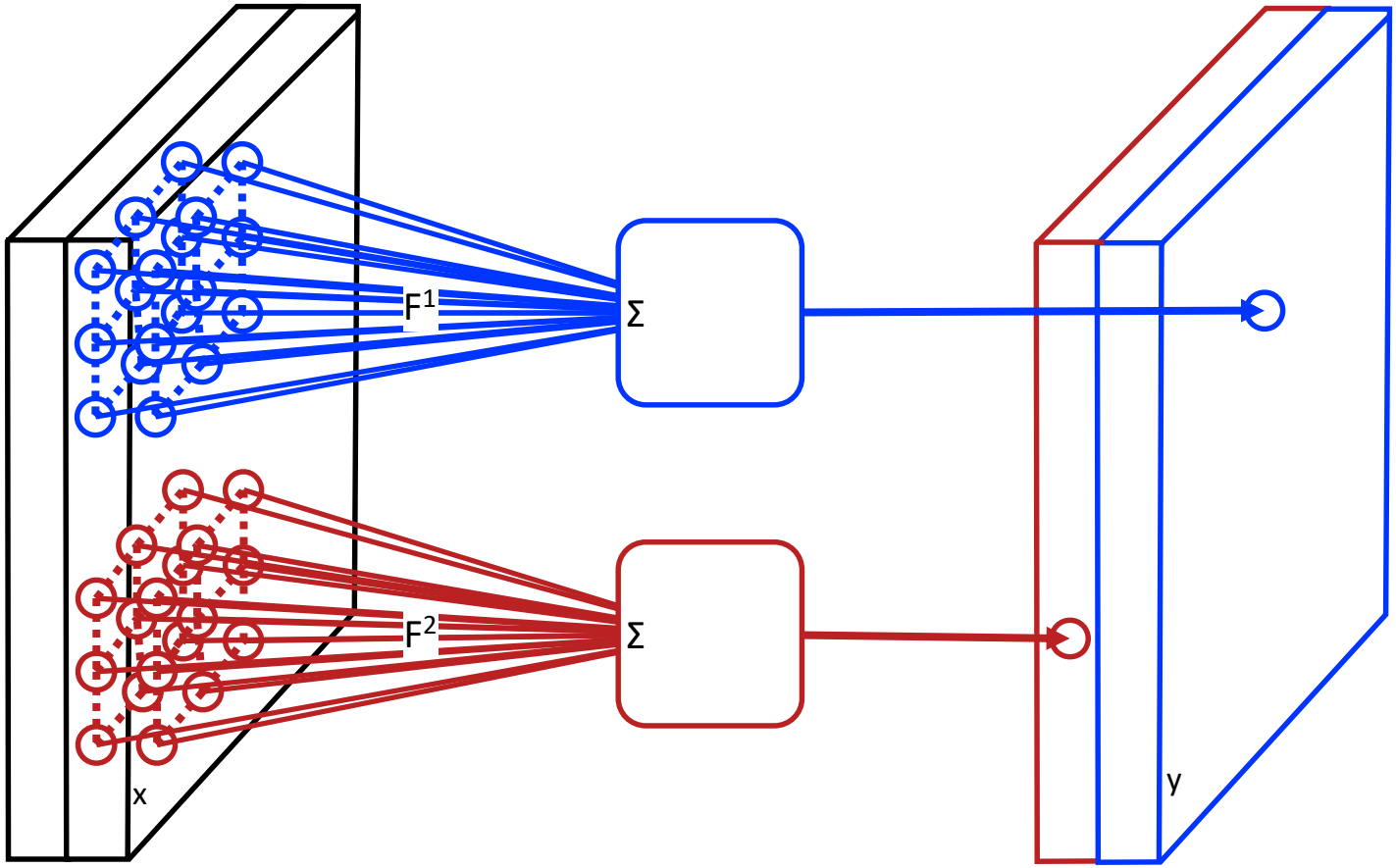
Convolutional layer

Why is it convolutional?

input features

a bank of 2 filters

2-dimensional output features

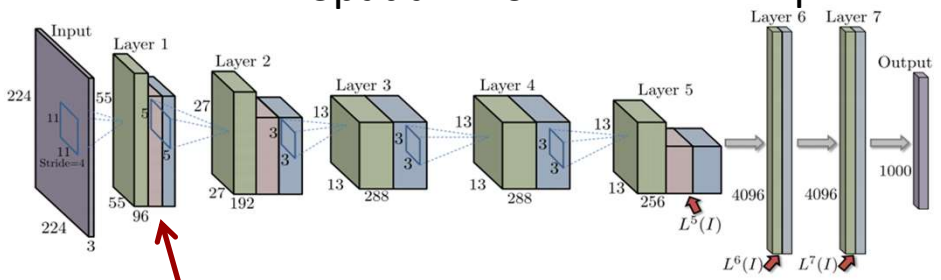


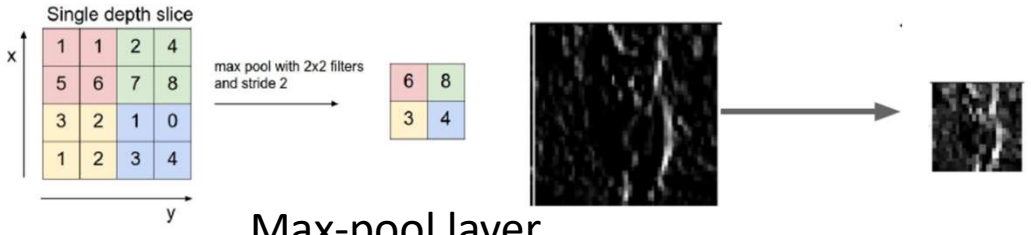
Convolution

Convolutional and Max-pooling layer

Spatial info

No spatial info



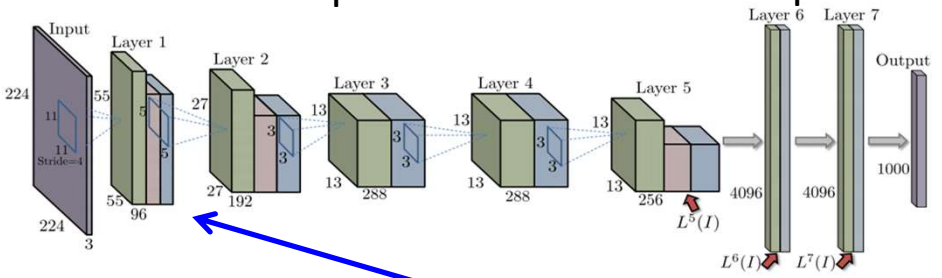


Max-pool layer

Convolutional and Max-pooling layer

Spatial info

No spatial info

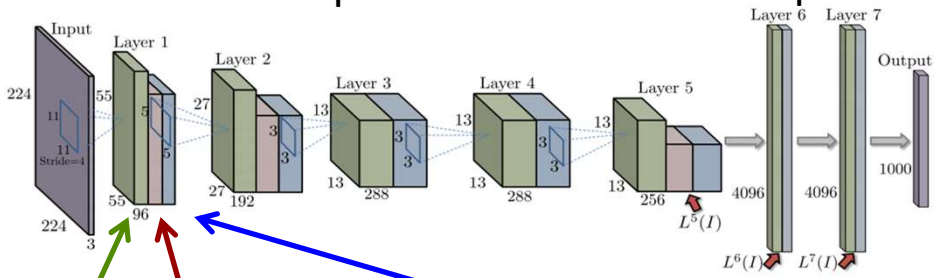


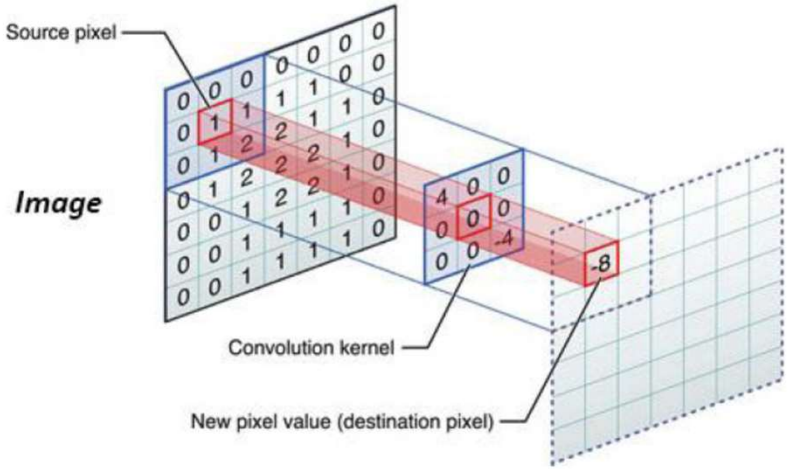
ReLU max(0,x)

Convolutional and Max-pooling layer

Spatial info

No spatial info

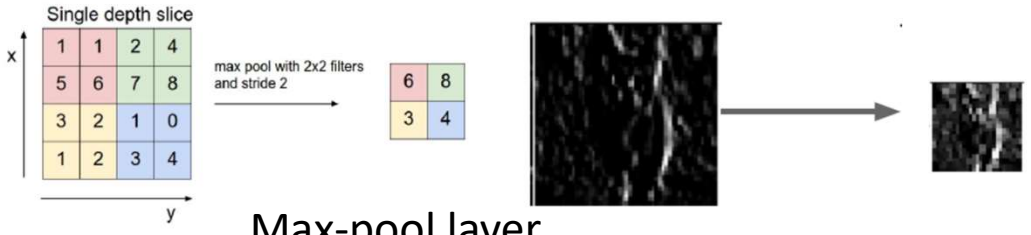




Filter Response

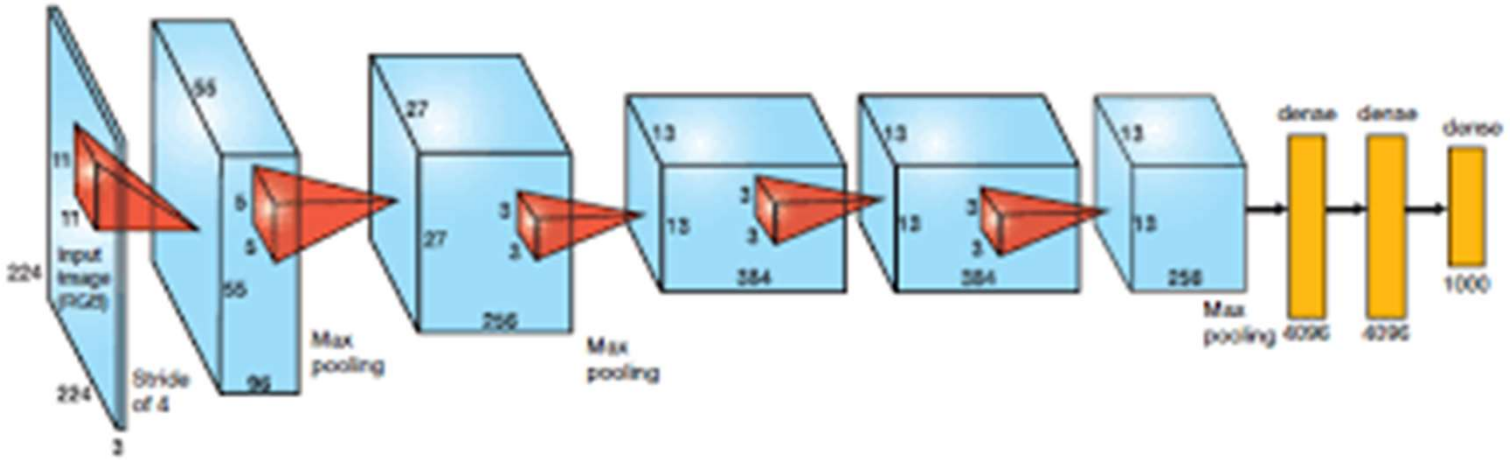
ReLU max(0,x)

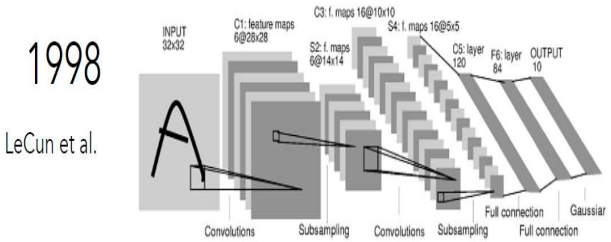
Convolutional layer



Max-pool layer

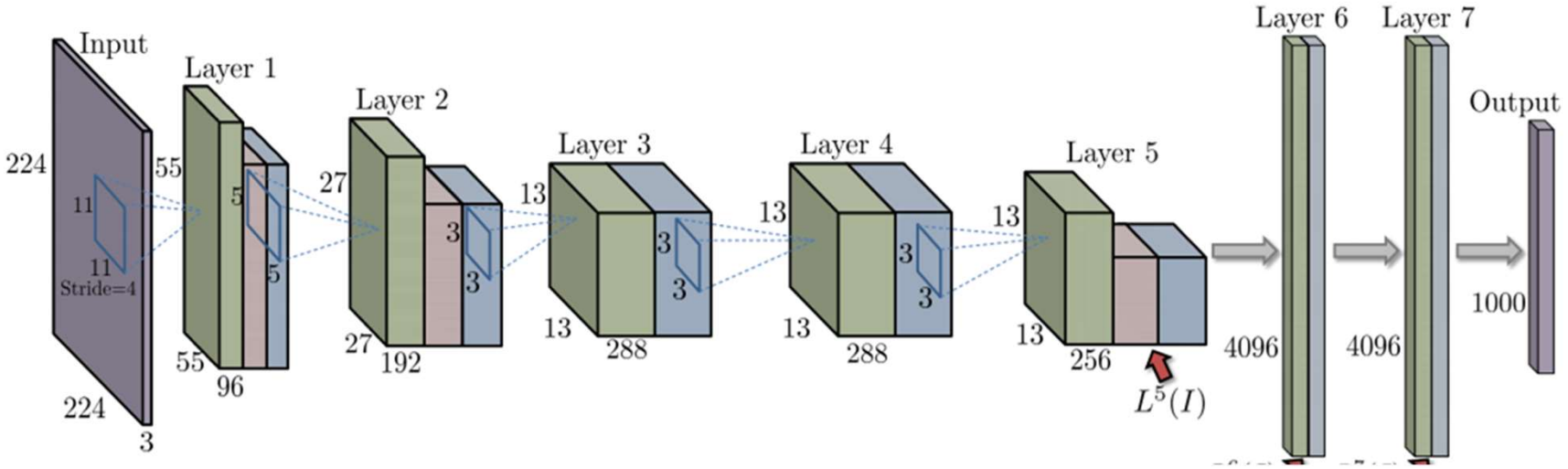
Why is it convolutional?





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

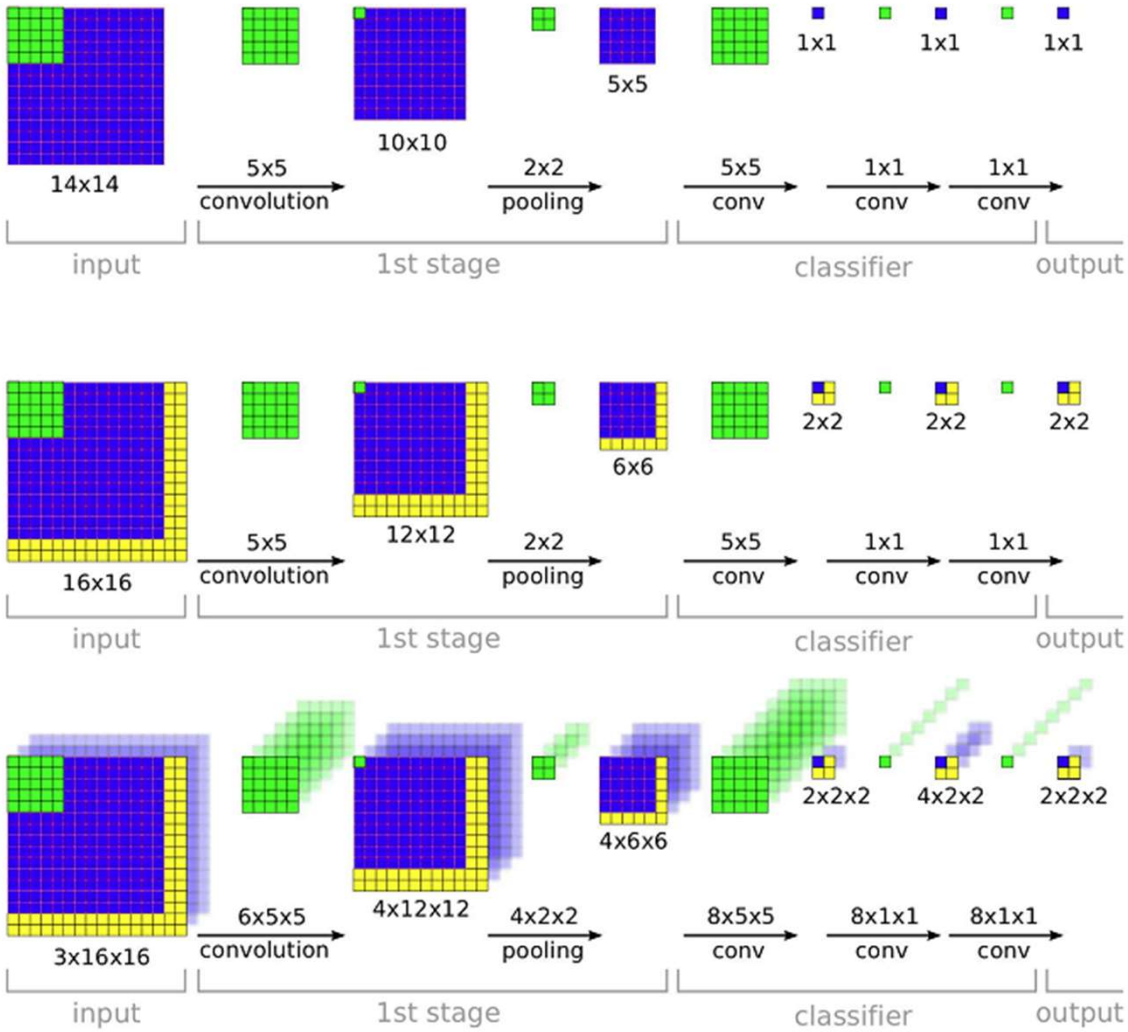
Training a CNN



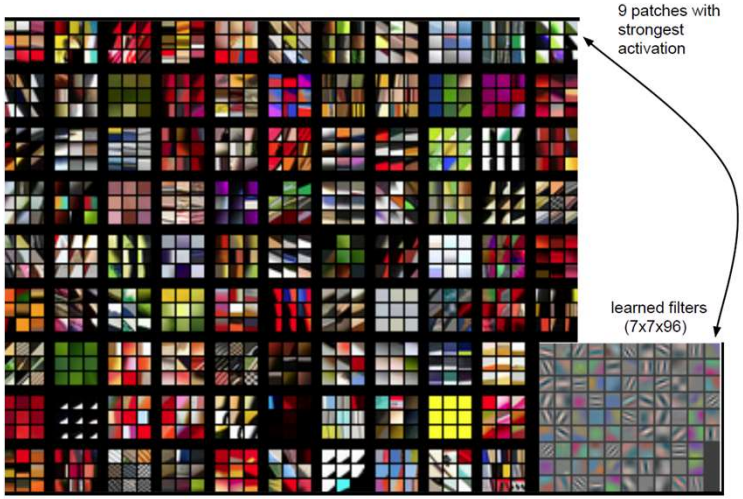
The process of training a CNN consists of training all hyperparameters: convolutional matrices and weights of the fully connected layers.

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

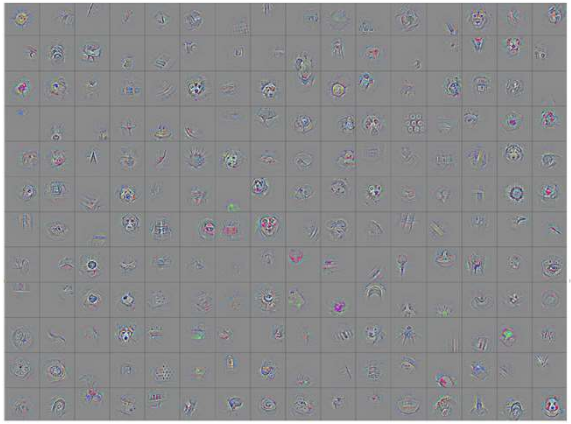
How does the CNN work?



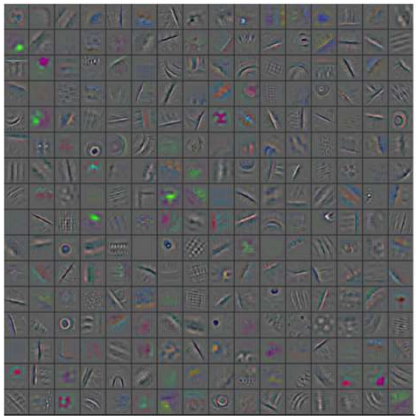
Learned convolutional filters



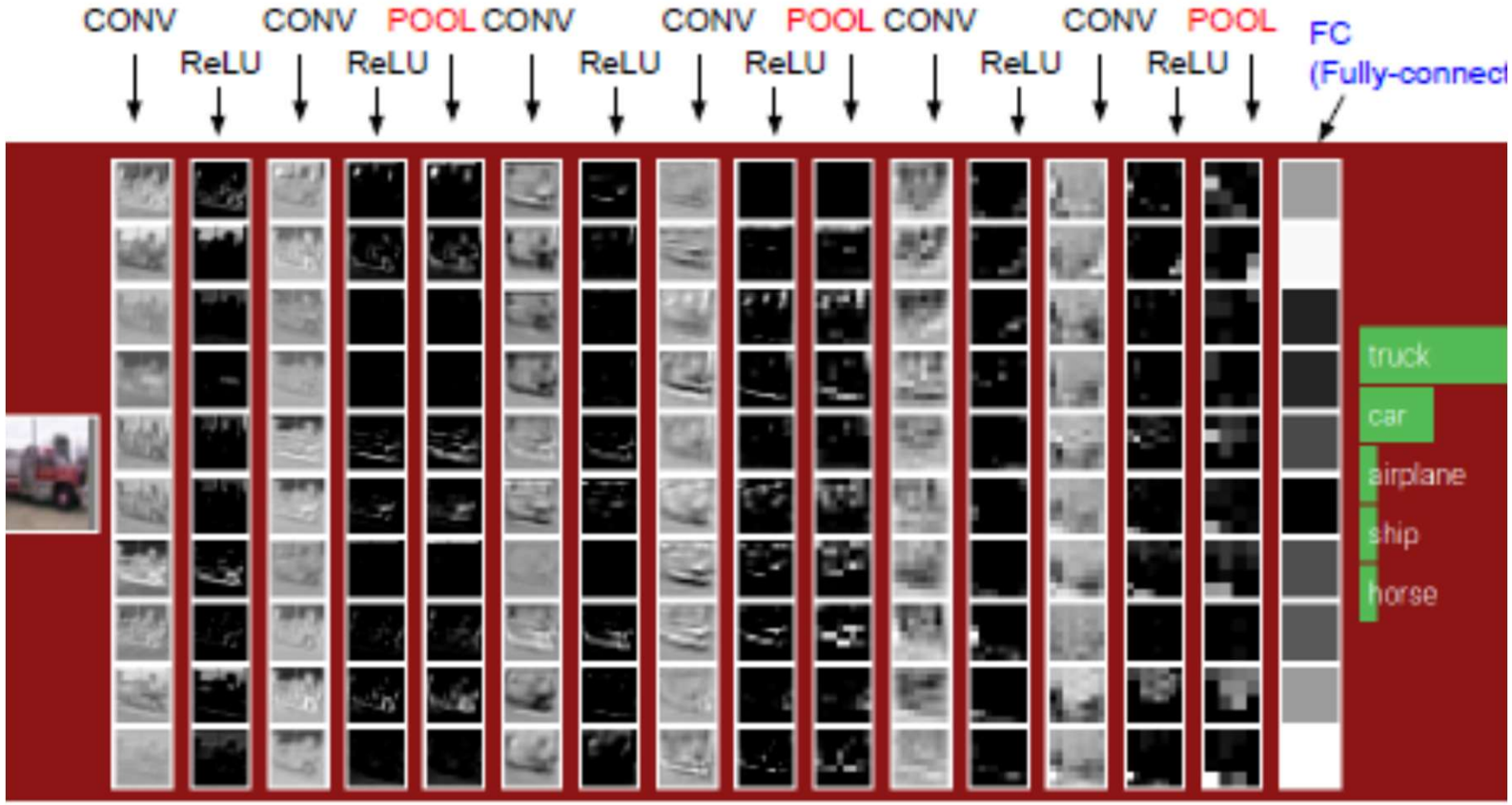
Strongest activations: Stage 5



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

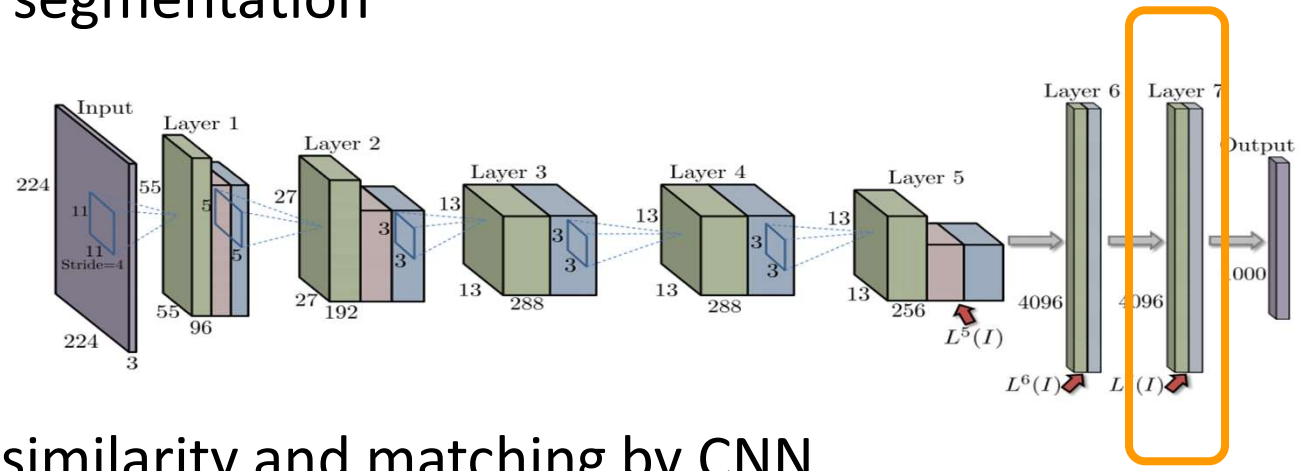


Image similarity and matching by CNN

Convolutional Neural Networks (4096 Features)

IMAGENET Large Scale Visual Recognition Challenge

Year 2010

NEC-UIUC

Dense descriptor grid: HOG, LBP

Coding: local coordinate, super-vector

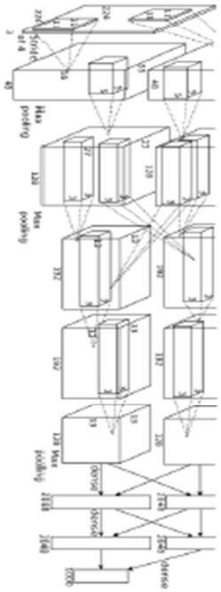
Pooling, SPM

Linear SVM

[Lin CVPR 2011]

Year 2012

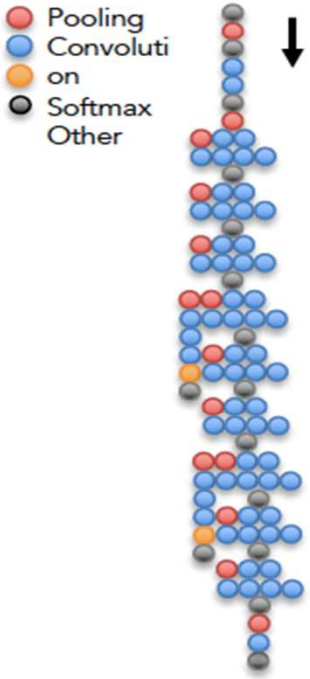
SuperVision



[Krizhevsky NIPS 2012]

Year 2014

GoogLeNet



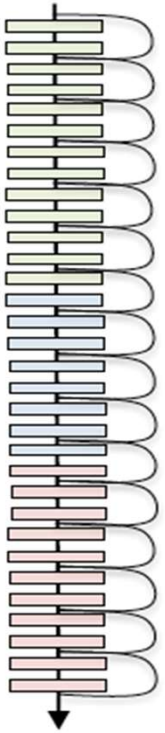
[Szegedy arxiv 2014]

VGG

[Simonyan arxiv 2014]

Year 2015

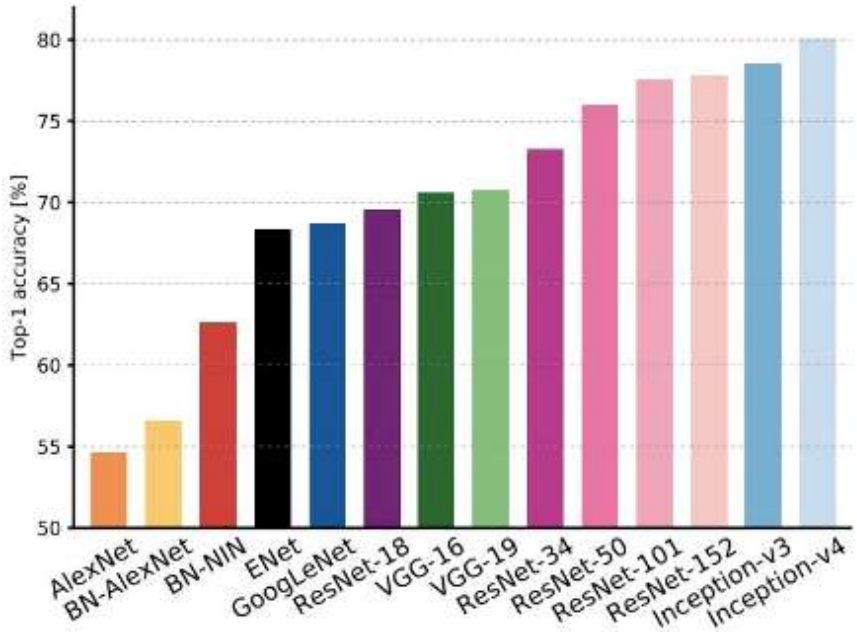
MSRA

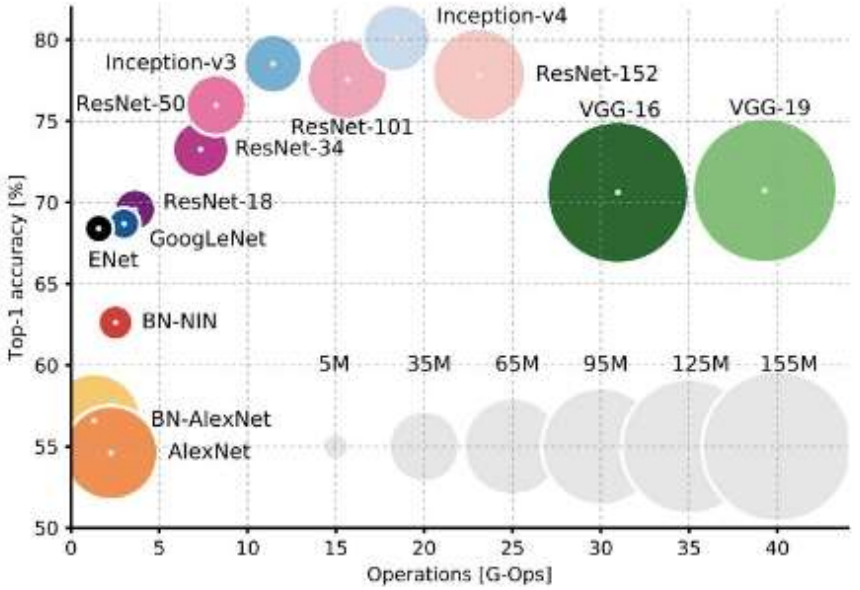


[He ICCV 2015]

$$o = f_1(f_2(f_3(\ldots . fn(x, \theta_n))))$$

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

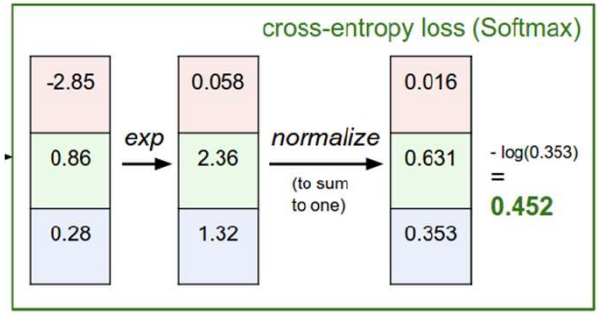
- AI, Machine learning & Deep learning
- What is a Convolutional Neural Network?
  - Layers
  - Loss function and CNN Optimization
- Applications

Loss function and optimisation

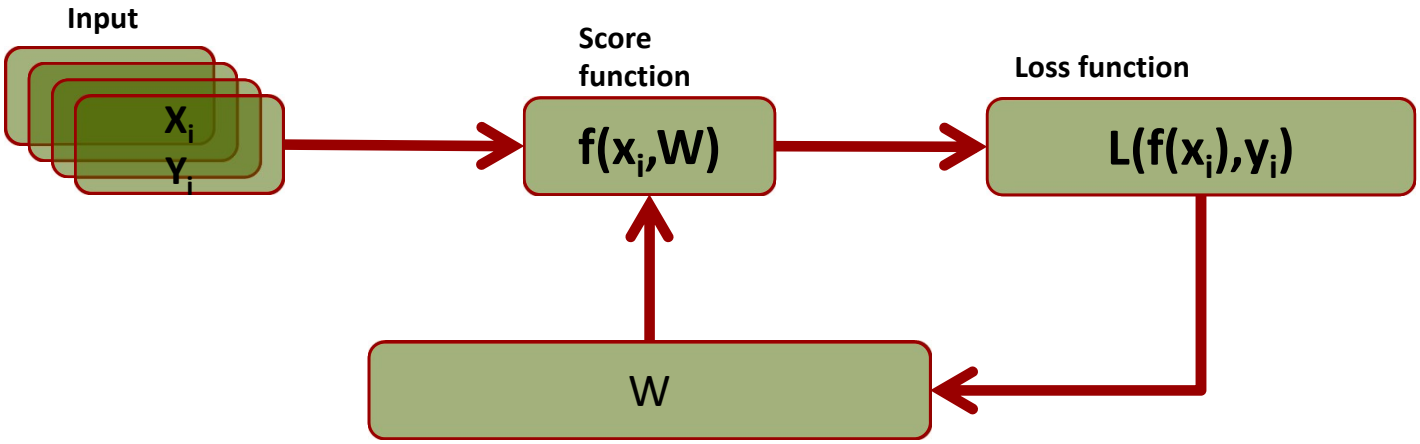
Question: if you were to assign a single number to how unhappy you are with these scores, what would you do?

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

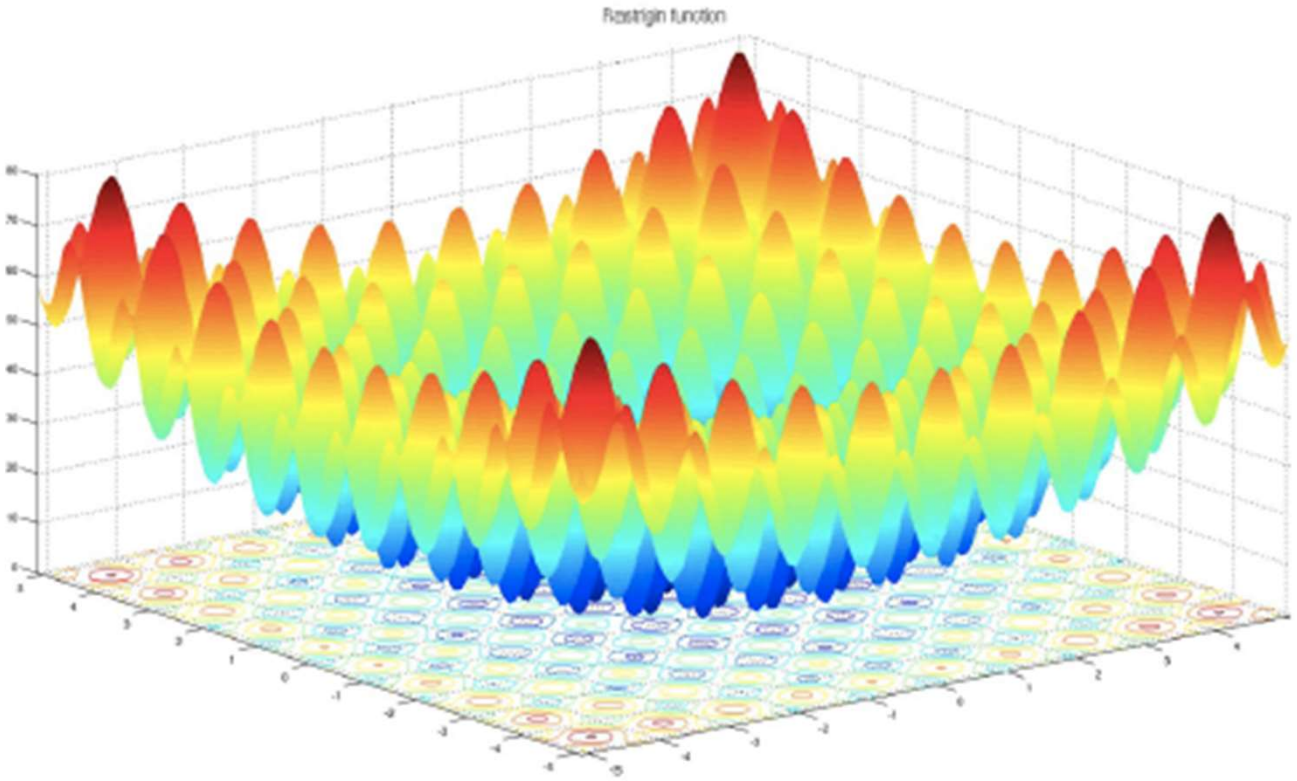
softmax function



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



Optimization

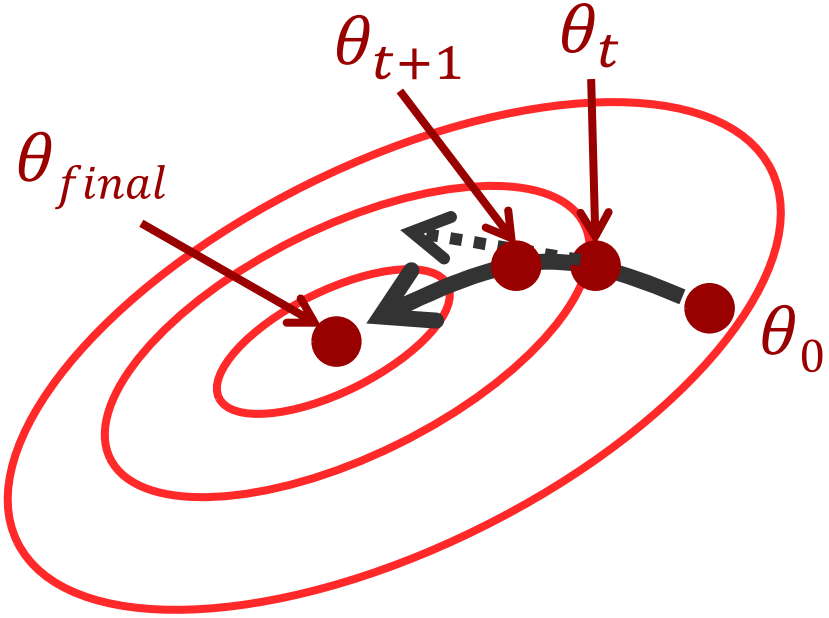


The process of training a CNN consists of training all hyperparameters: convolutional matrices and weights of the fully connected layers.

- **Millions of parameters!!!**

Gradient descent

$$o = f_1(f_2(f_3(\ldots . fn(x, \theta_n))))$$



- Initialize θ_0 randomly
- For t in 0,…, T_{maxiter}

$$\theta^{t+1} = \theta^t - \eta_t \cdot \nabla L(\theta^t)$$

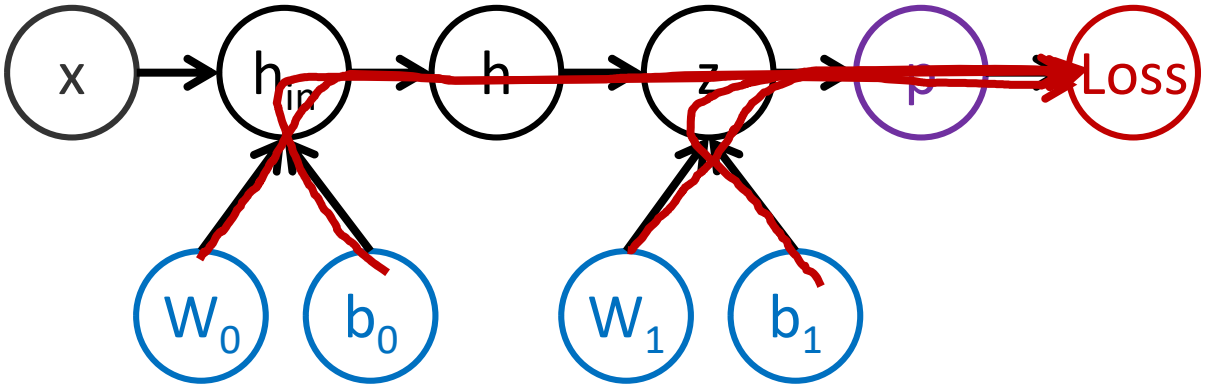
Gradient of the objective

Learning rate (step size)

- Computation of $\nabla L(\theta^t)$ requires a full sweep over the training data
- Per-iteration comp. cost = O(n)

Chain rule

- Identify how each variable influence the loss



$$\frac{\partial \text{Loss}}{\partial W_1} = \frac{\partial Loss}{\partial p} \cdot \cdots \cdot \frac{\partial z}{\partial W_1}$$

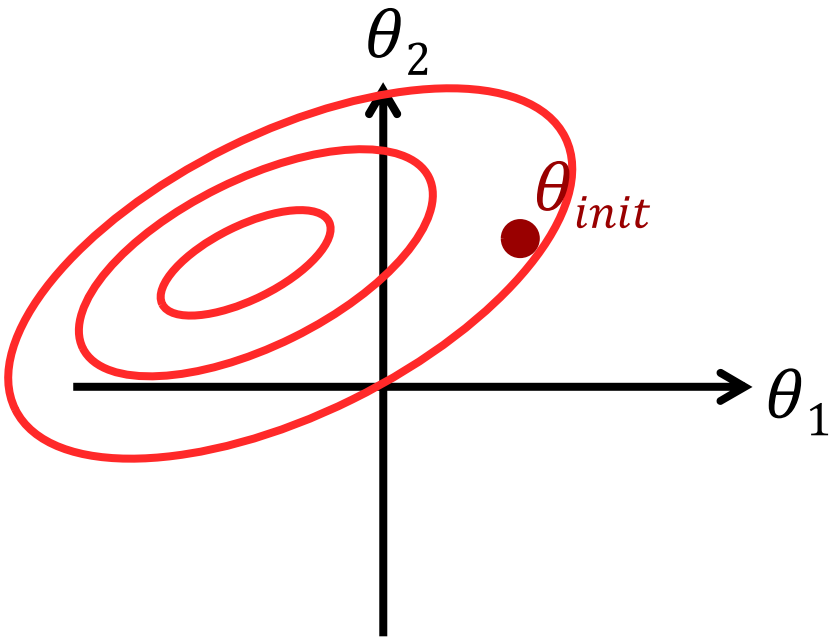
$$\frac{\partial \text{Loss}}{\partial b_1} = \frac{\partial Loss}{\partial p} \cdot \cdots \cdot \frac{\partial z}{\partial b_1}$$

$$\frac{\partial \text{Loss}}{\partial W_0} = \frac{\partial Loss}{\partial p} \cdot \cdots \cdot \frac{\partial h_{in}}{\partial W_0}$$

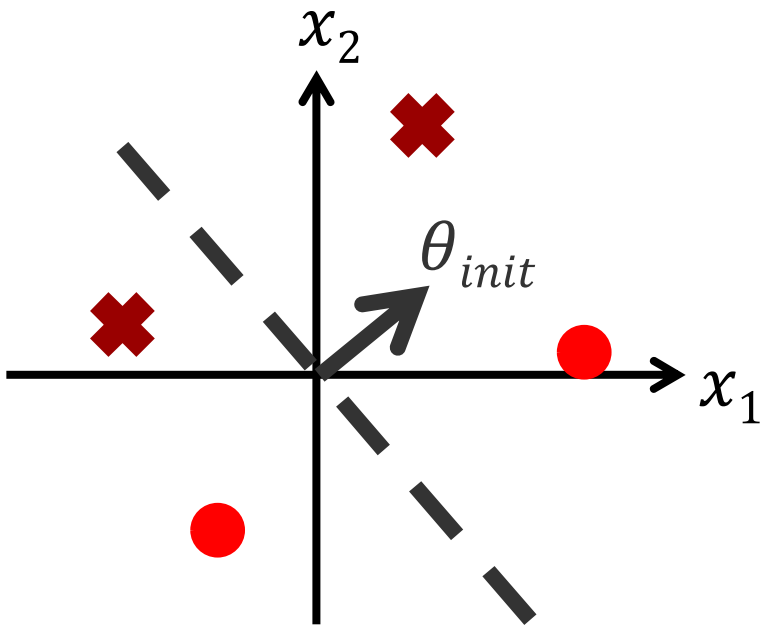
$$\frac{\partial \text{Loss}}{\partial b_0} = \frac{\partial Loss}{\partial p} \cdot \cdots \cdot \frac{\partial h_{in}}{\partial b_0}$$

Landscape of training objective

Parameter space

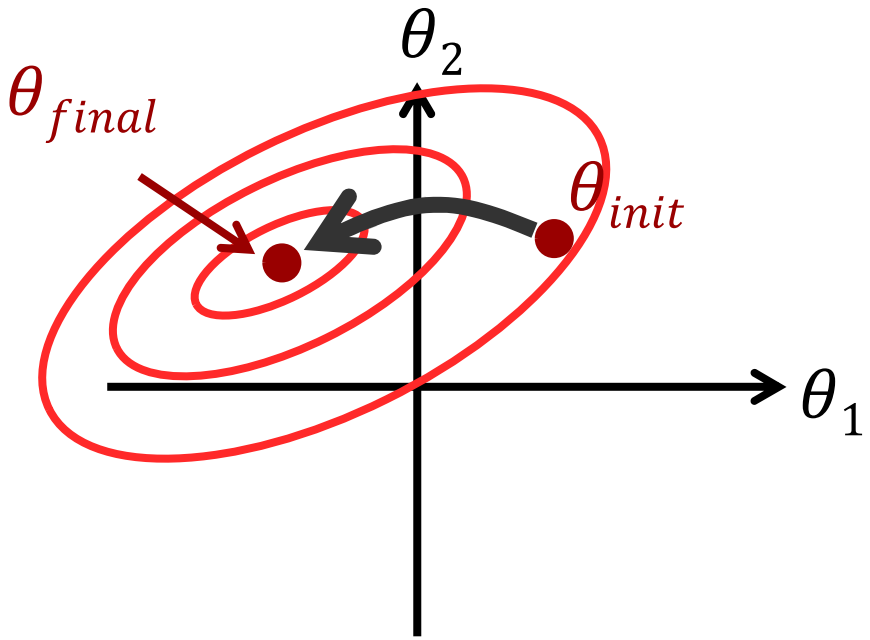


Example space

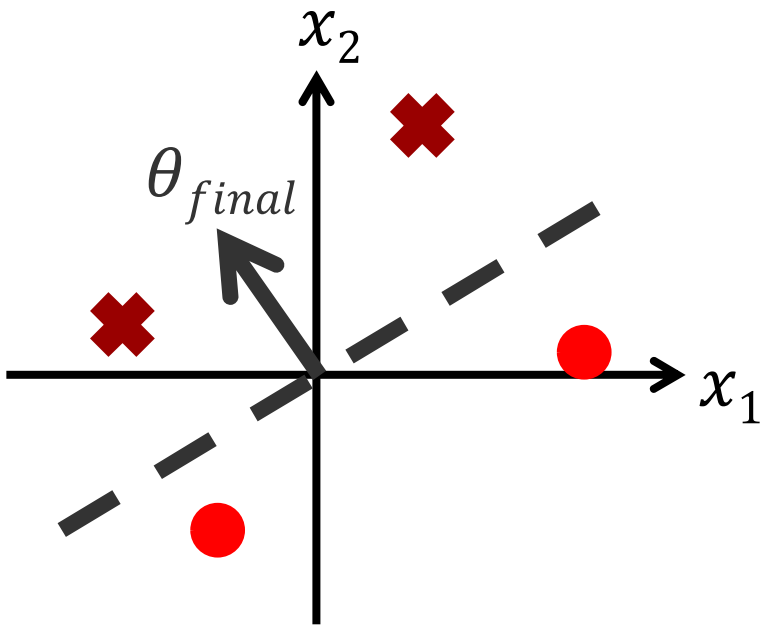


Landscape of training objective

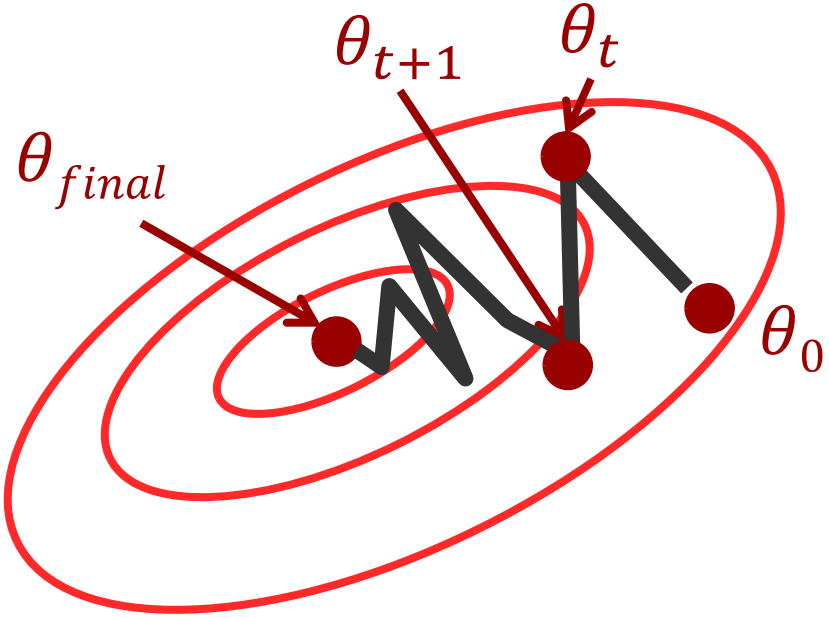
Parameter space



Example space



Stochastic gradient descent (SGD)



- Initialize θ_0 randomly
- For t in 0,…, T_{maxiter}

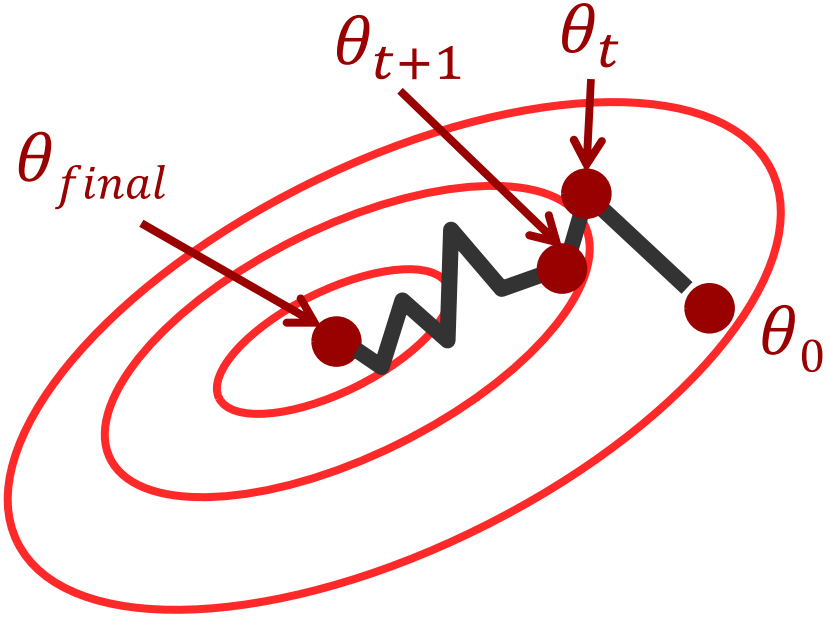
$$\theta^{t+1} = \theta^t - \eta_t \cdot \nabla Loss(f_\theta(x_i), y_i)$$

Stochastic gradient

where index *i* is chosen randomly

- computation of $\nabla Loss(\dots)$ requires only one training example
- Per-iteration comp. cost = O(1)

Mini-batch stochastic gradient descent



- Initialize θ_0 randomly
- For t in 0,…, T_{maxiter}

$$\theta^{t+1} = \theta^t - \eta_t \cdot \tilde{\nabla}_B L(\theta)$$

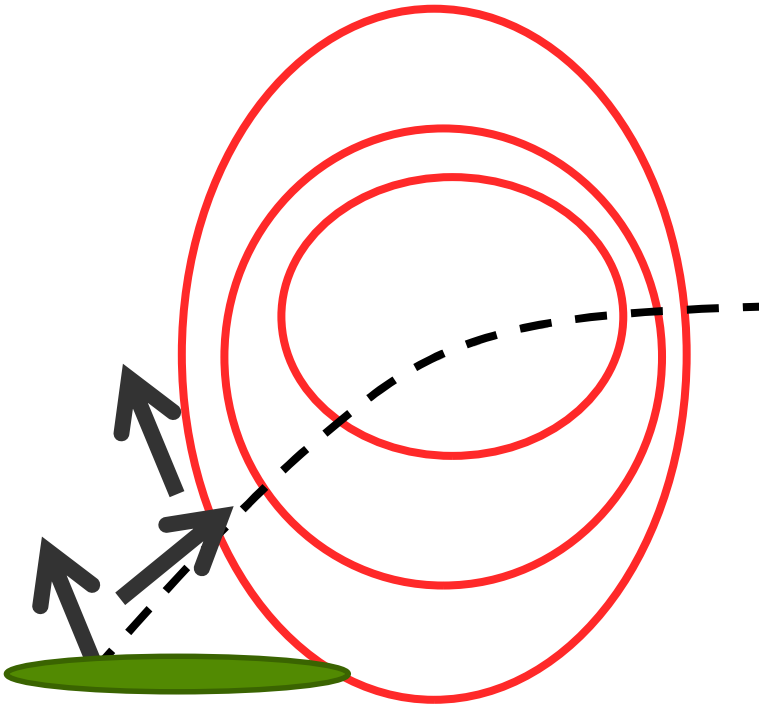
minibatch gradient

where minibatch *B* is chosen randomly

- $\tilde{\nabla} L(\theta)$ is average gradient over random subset of data of size *B*
- Per-iteration comp. cost = O(*B*)

More optimization algorithms

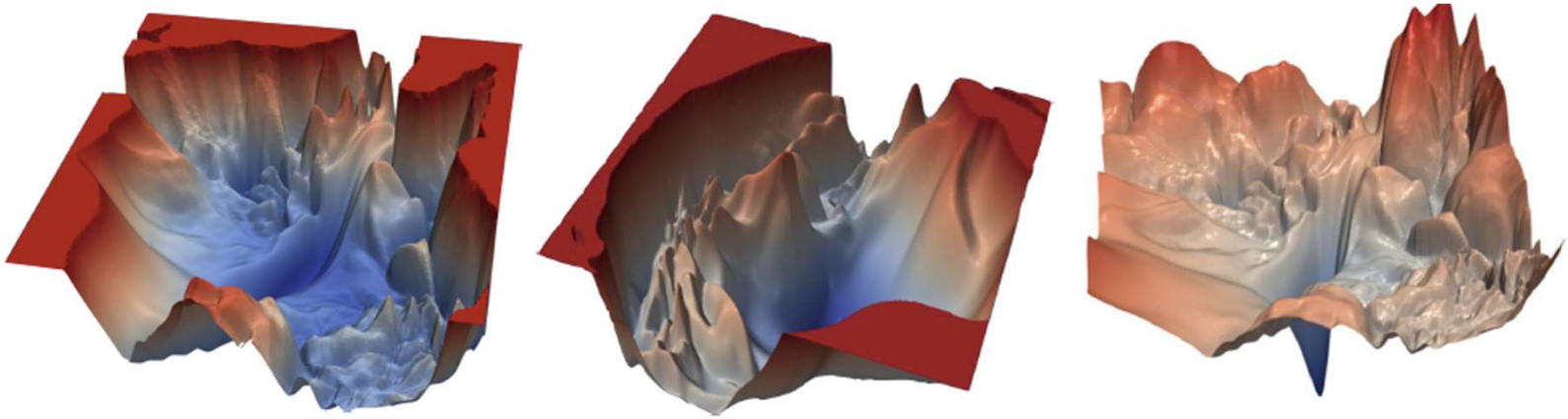
- **Momentum SGD**: improves SGD by incorporating “momentum”



- **Adam** [Kingma & Ba 2015]: uses first and second order statistics of the gradients so that gradients are normalized

- **Benefit**: prevents the vanishing/exploding gradient problem

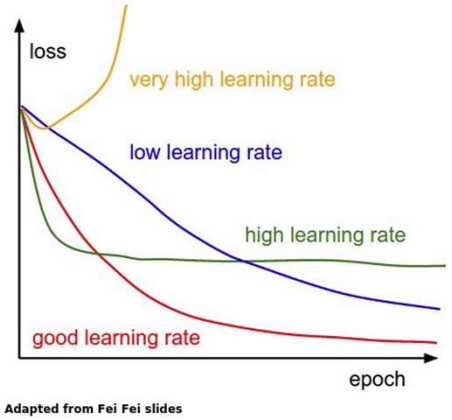
Stochastic Gradient Descent



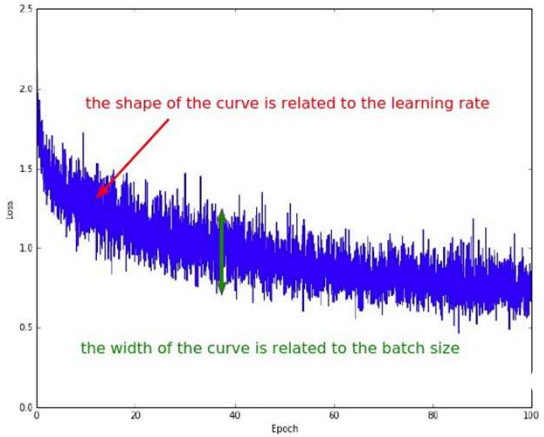
VGG-56 VGG-110 ResNet-56

Hao Li et al., NIPS, 2017

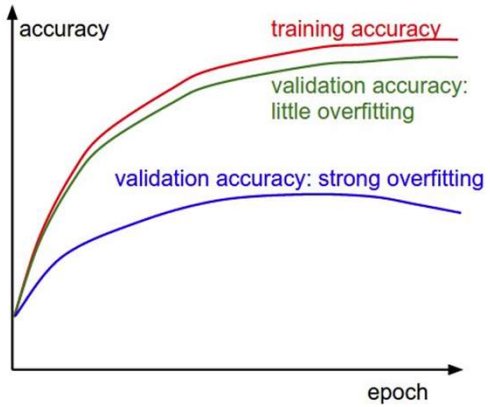
Monitoring loss and accuracy

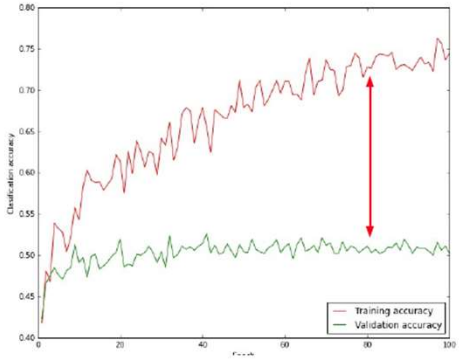


Adapted from Fei Fei slides



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

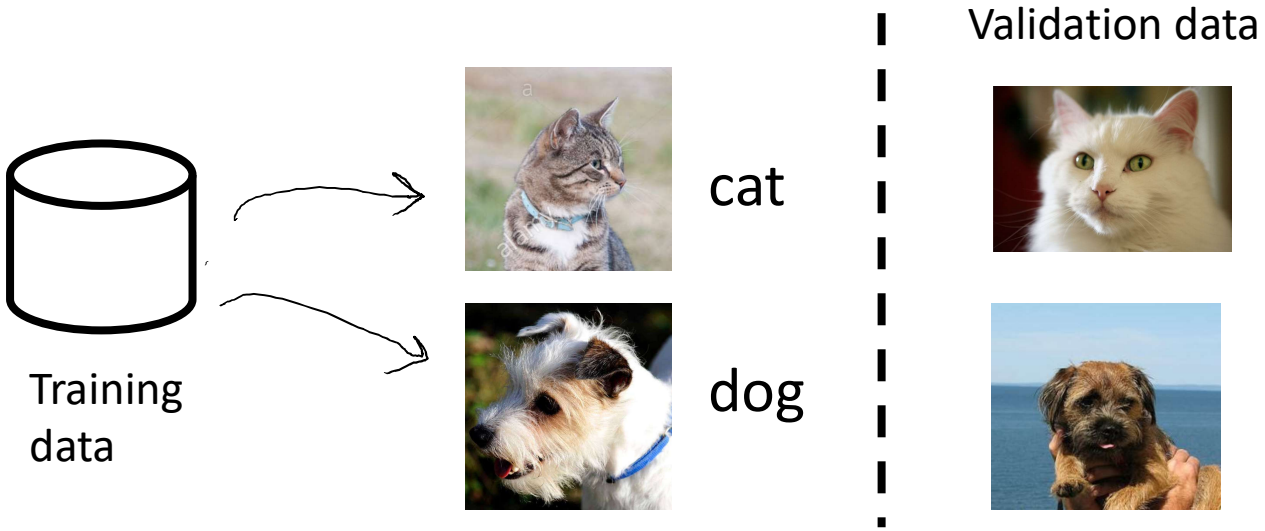




Big gap?
- you're overfitting, increase regularization!

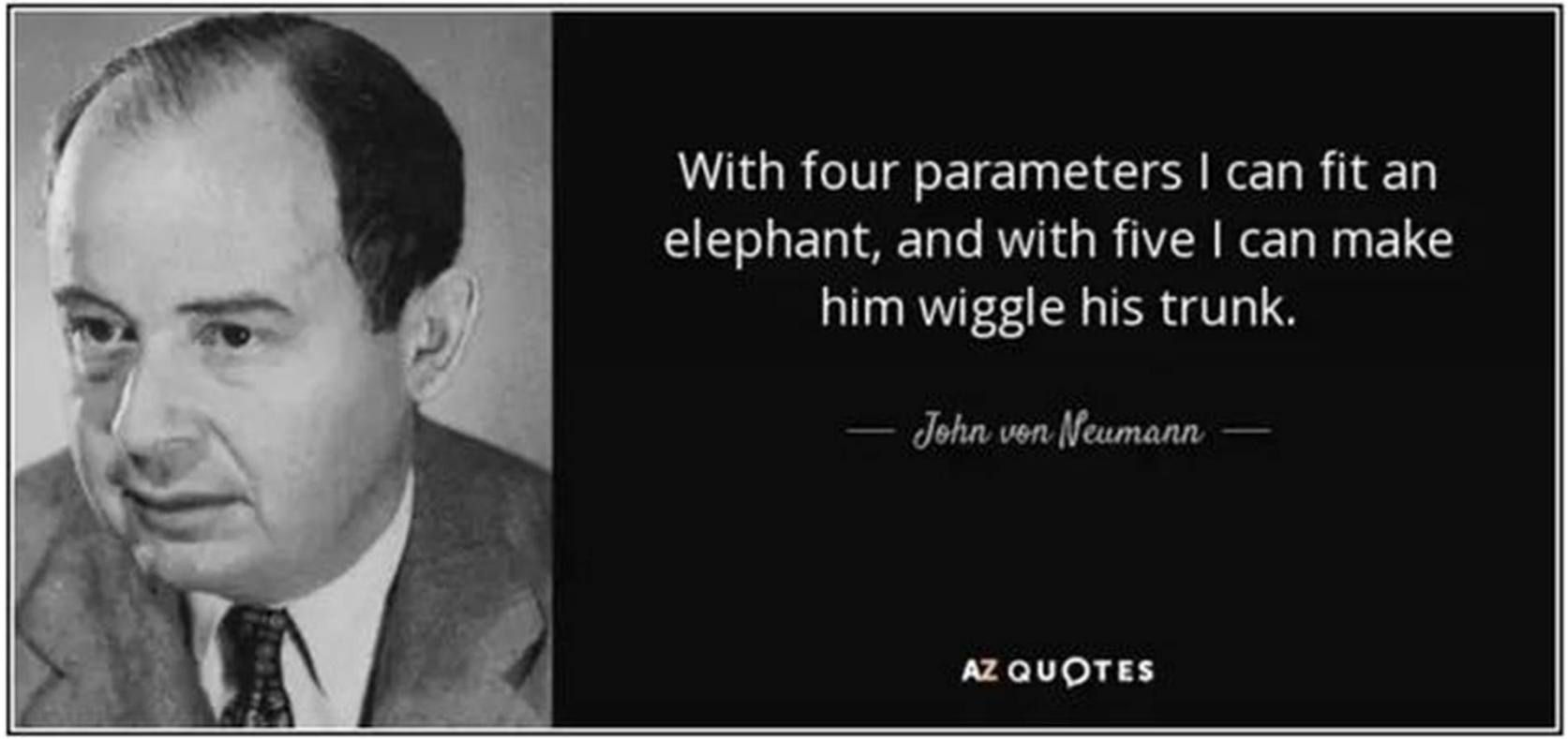
Overfitting – what is signal vs noise?

- Imagine:



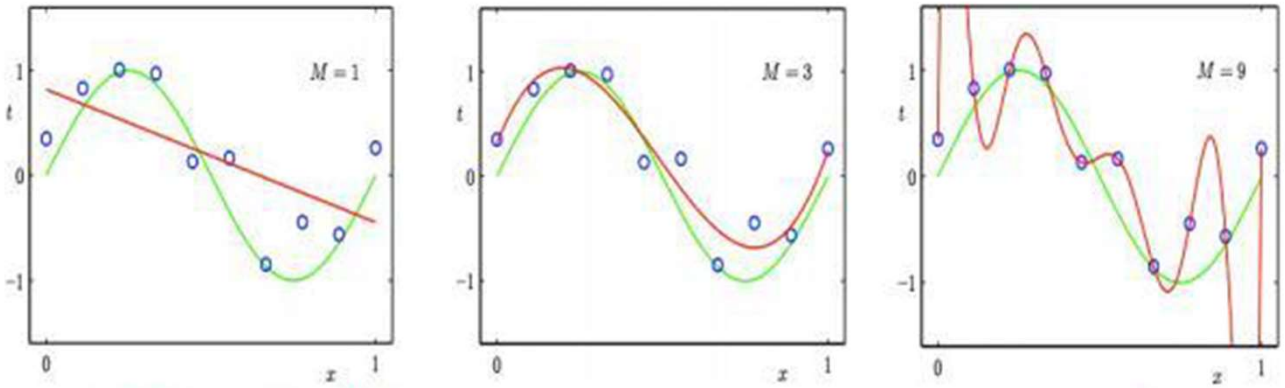
- Powerful models are more likely to overfit
- We need validation data: leave out some portion of the training data to validate the generalizability of the model

Overfitting



Under- and Over-fitting

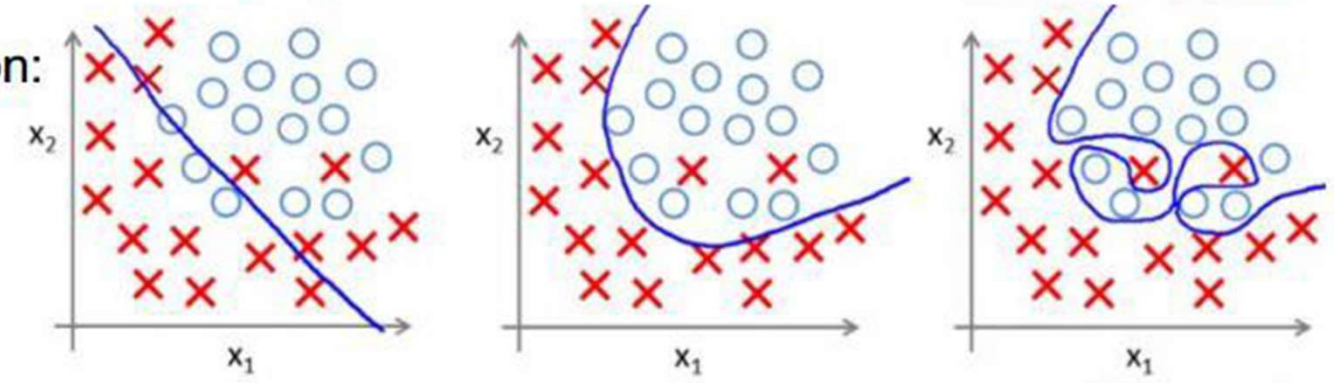
Regression:



predictor too inflexible:
cannot capture pattern

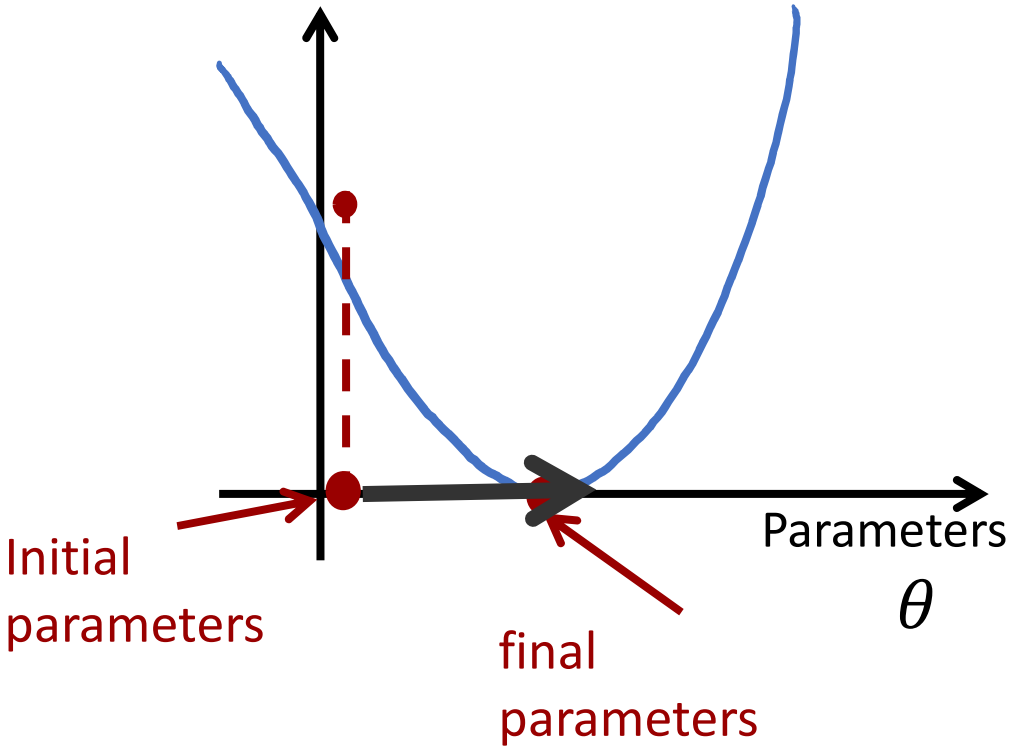
predictor too flexible:
fits noise in the data

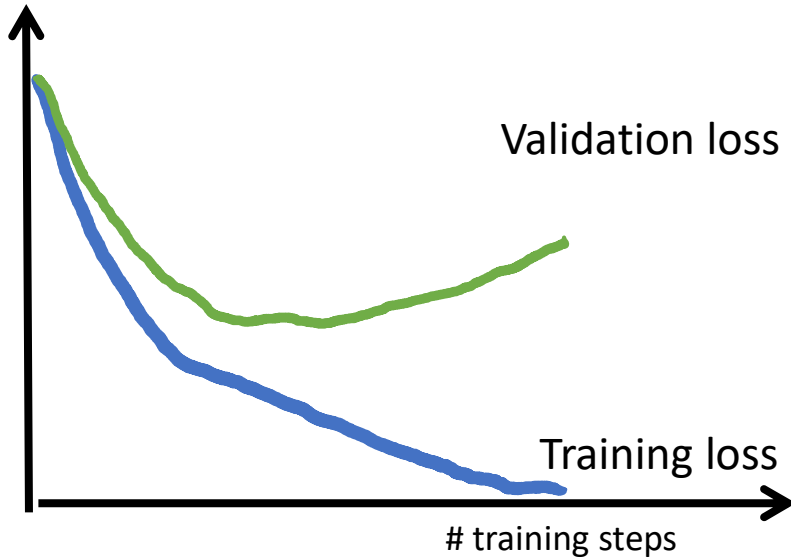
Classification:



Landscape of training objective

Objective function: $L(\theta) = \frac{1}{n}\sum_{i=1}^{n} Loss(f_\theta(x_i), y_i)$

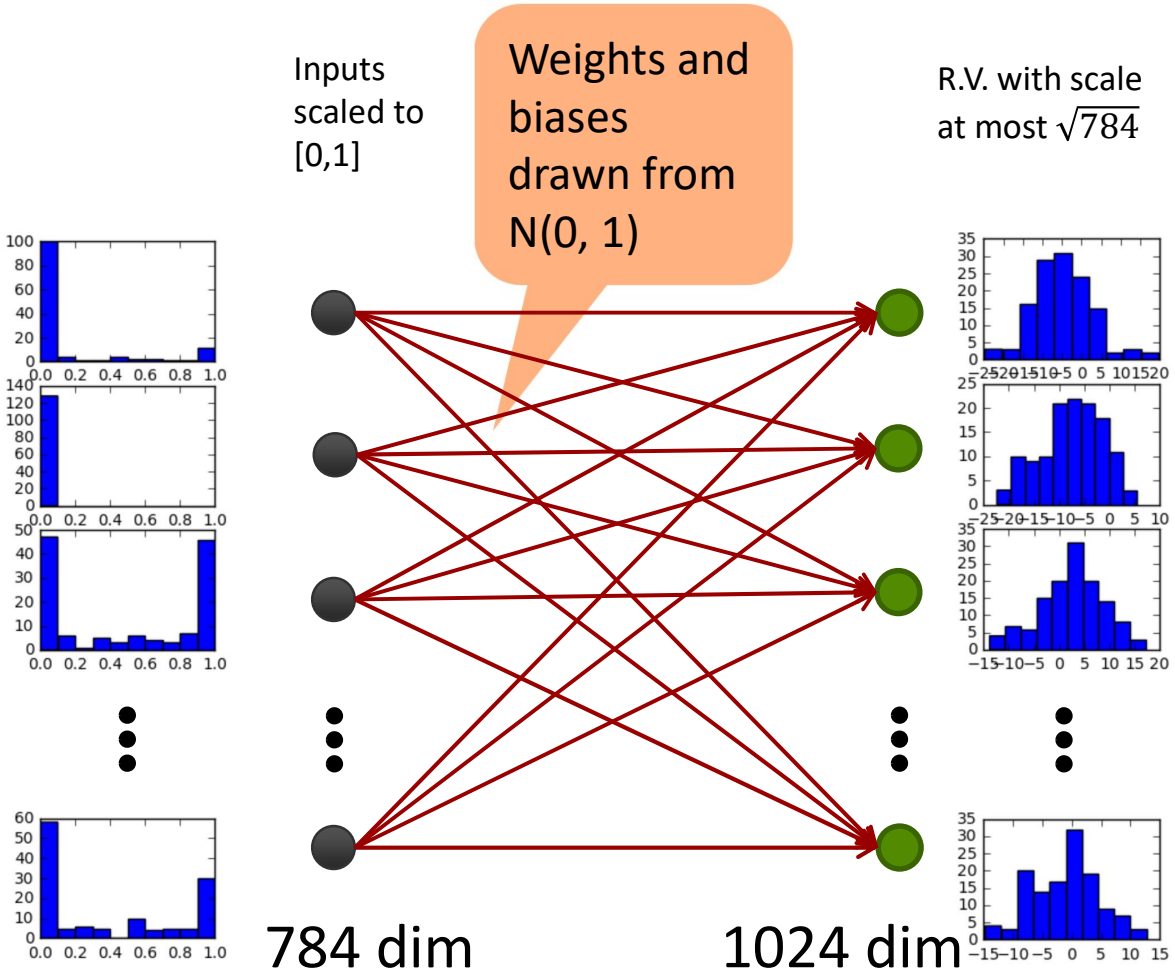




Techniques to reduce overfitting

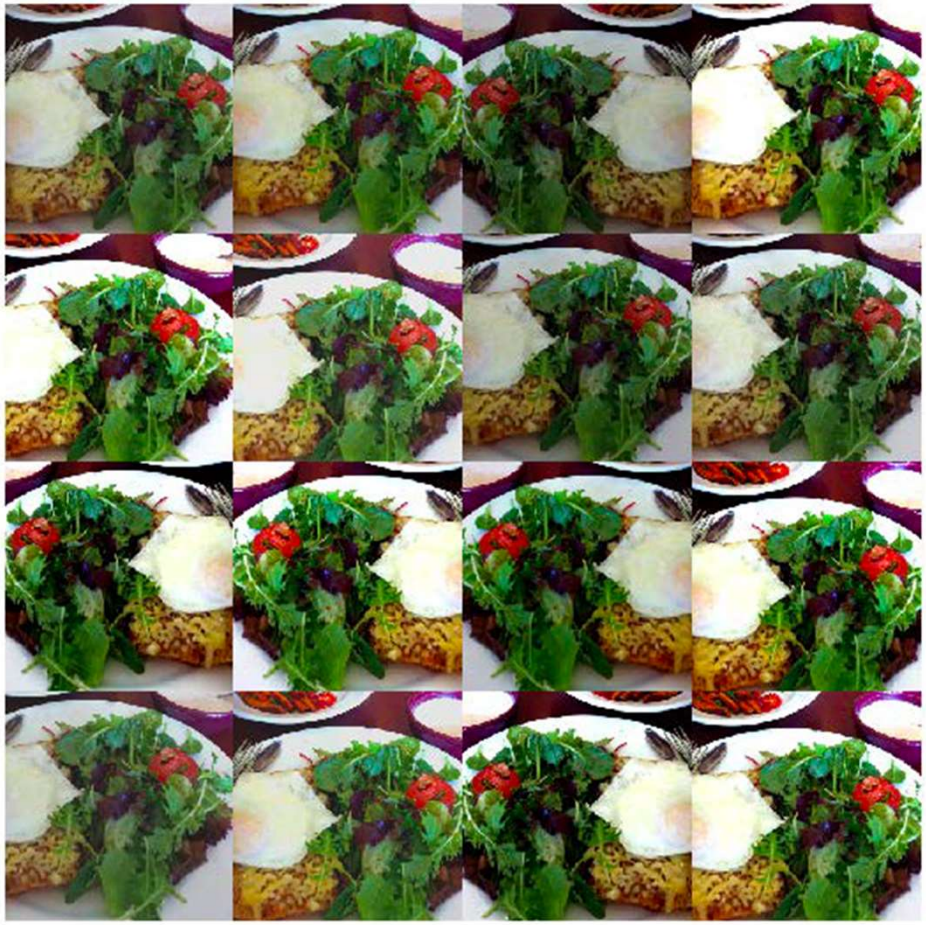
- Reduce the number of parameters
  - Parameter sharing (convnets, recurrent neural nets)
- Weight decay (aka L2 regularization)
  - Penalizes the magnitude of the parameters $\sum_{j=1}^{d} w_j^2$
- Early stopping
  - Indirectly controls the magnitude of the parameters
- Alternatives
  - Dropout, batch normalization

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

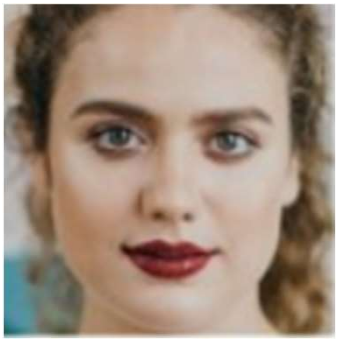
Everybody dances now

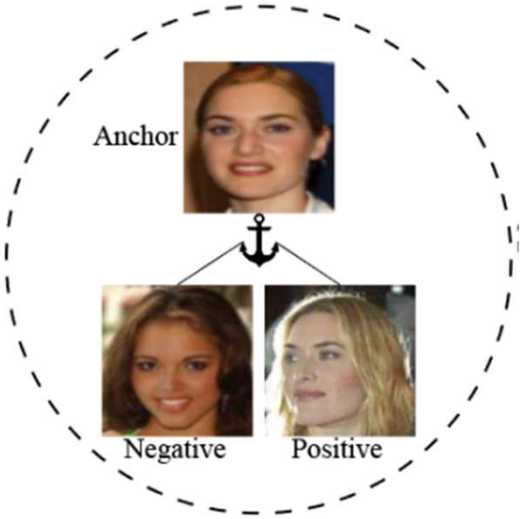
More links

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)

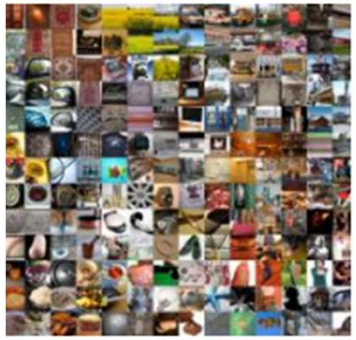
Shopping (Mango, Amazon)

Biometric data for cheating detection

Analizes up to 38 faciales facial micro-gestures, digital prints and scans hand veins

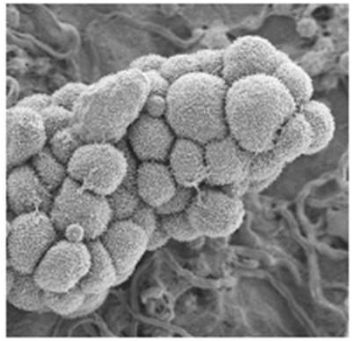
iBorderCtrl will reduce the cost of frontiers access of near 700M crossing European countries

Deep leearning everywhere



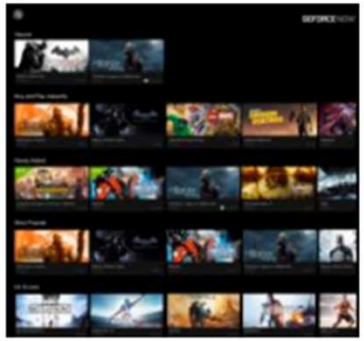
INTERNET & CLOUD

Image Classification
Speech Recognition
Language Translation
Language Processing
Sentiment Analysis
Recommendation



MEDICINE & BIOLOGY

Cancer Cell Detection
Diabetic Grading
Drug Discovery



MEDIA & ENTERTAINMENT

Video Captioning
Video Search
Real Time Translation

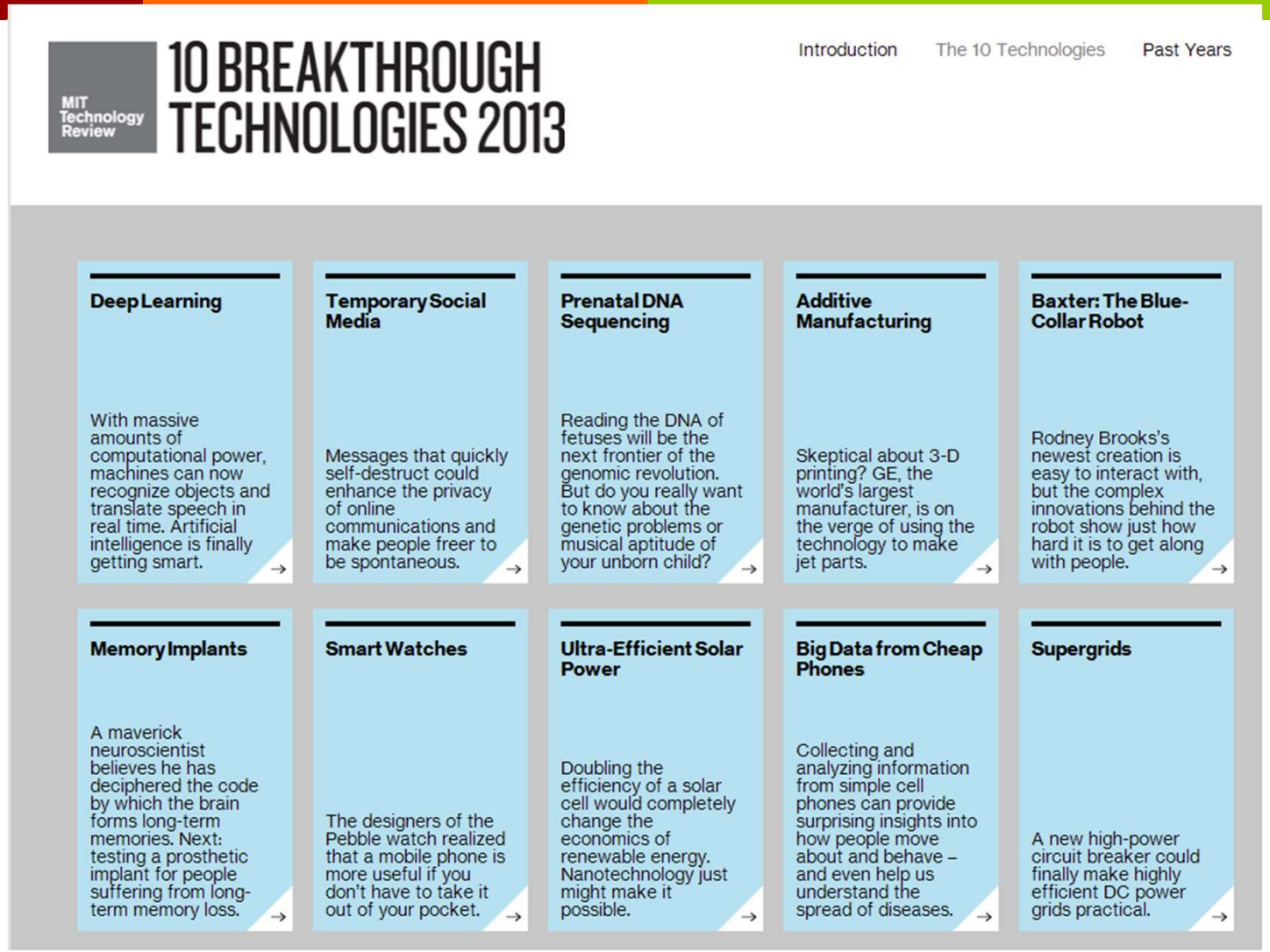
SECURITY & DEFENSE

Face Detection
Video Surveillance
Satellite Imagery

AUTONOMOUS MACHINES

Pedestrian Detection
Lane Tracking
Recognize Traffic Sign

Deep learning - one of the 10 breakthrough technologies



Deep leearning everywhere

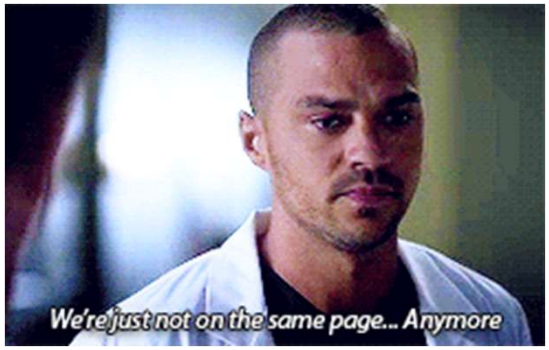
"Preparing for the Future of Artificial Intelligence." ★1

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/

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

CNN applicable to different Computer Vision problems

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