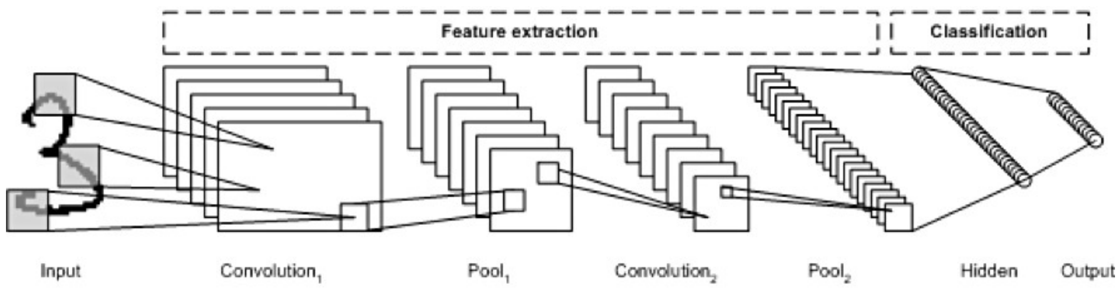
Convolutional Neural Networks



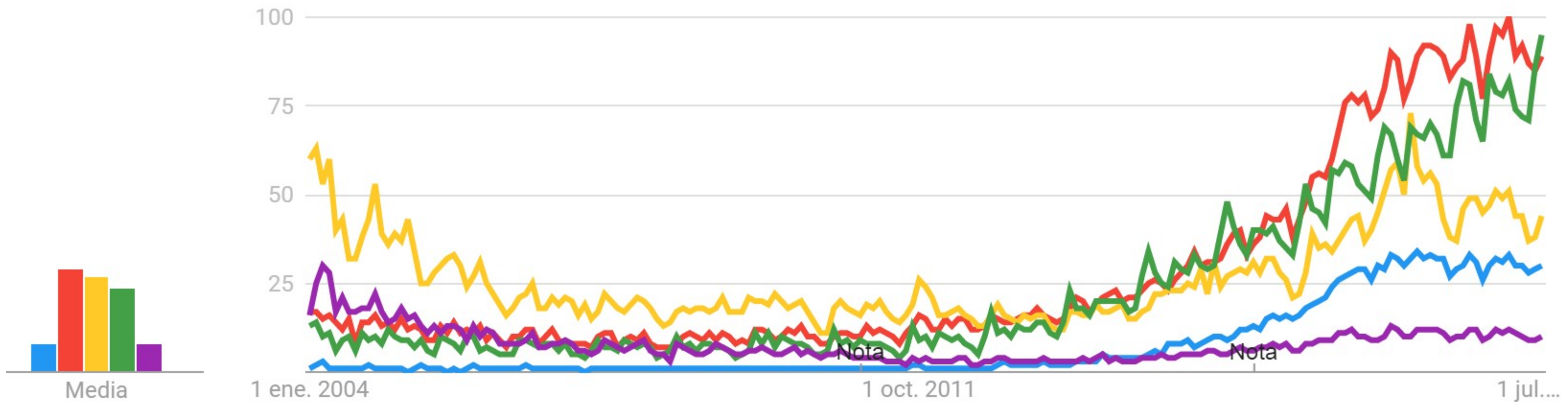
Class 7: Petia Radeva

Index

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

Artificial Intelligence

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



Google Scholar reveals its most influential papers


nature International weekly journal of science

Deep learning

Yann LeCun, Yoshua Bengio & Geoffrey Hinton

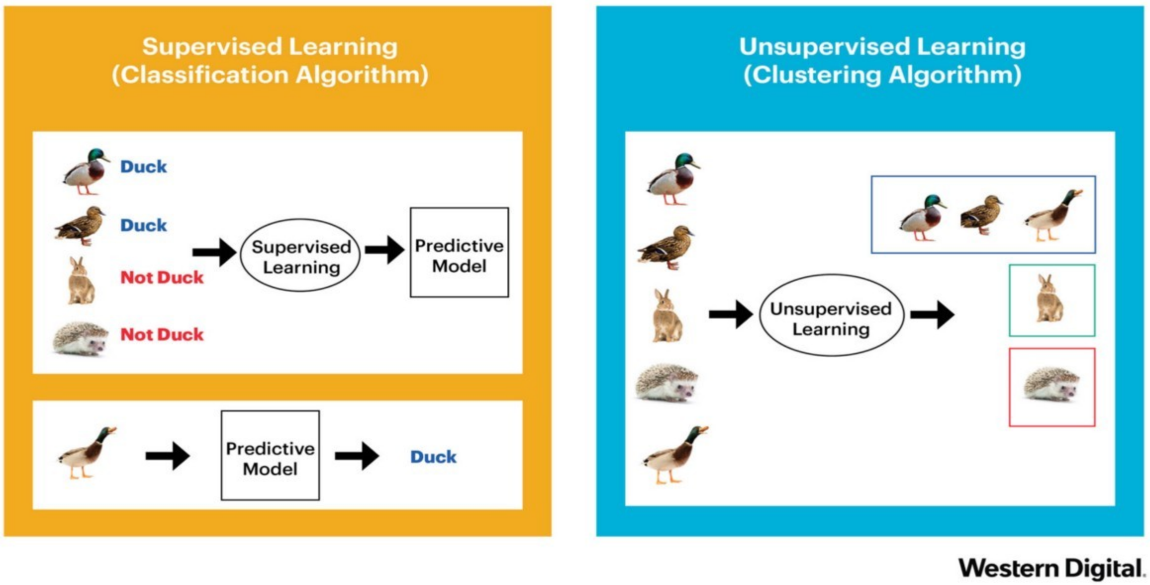
Affiliations | Corresponding author

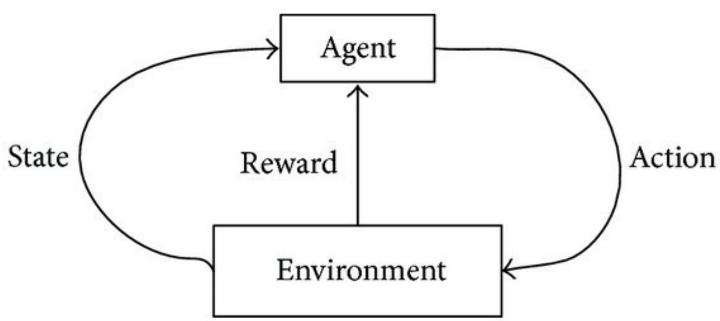
Nature 521, 436–444 (28 May 2015) | doi:10.1038/nature14539

Received 25 February 2015 | Accepted 01 May 2015 | Published online 27 May 2015

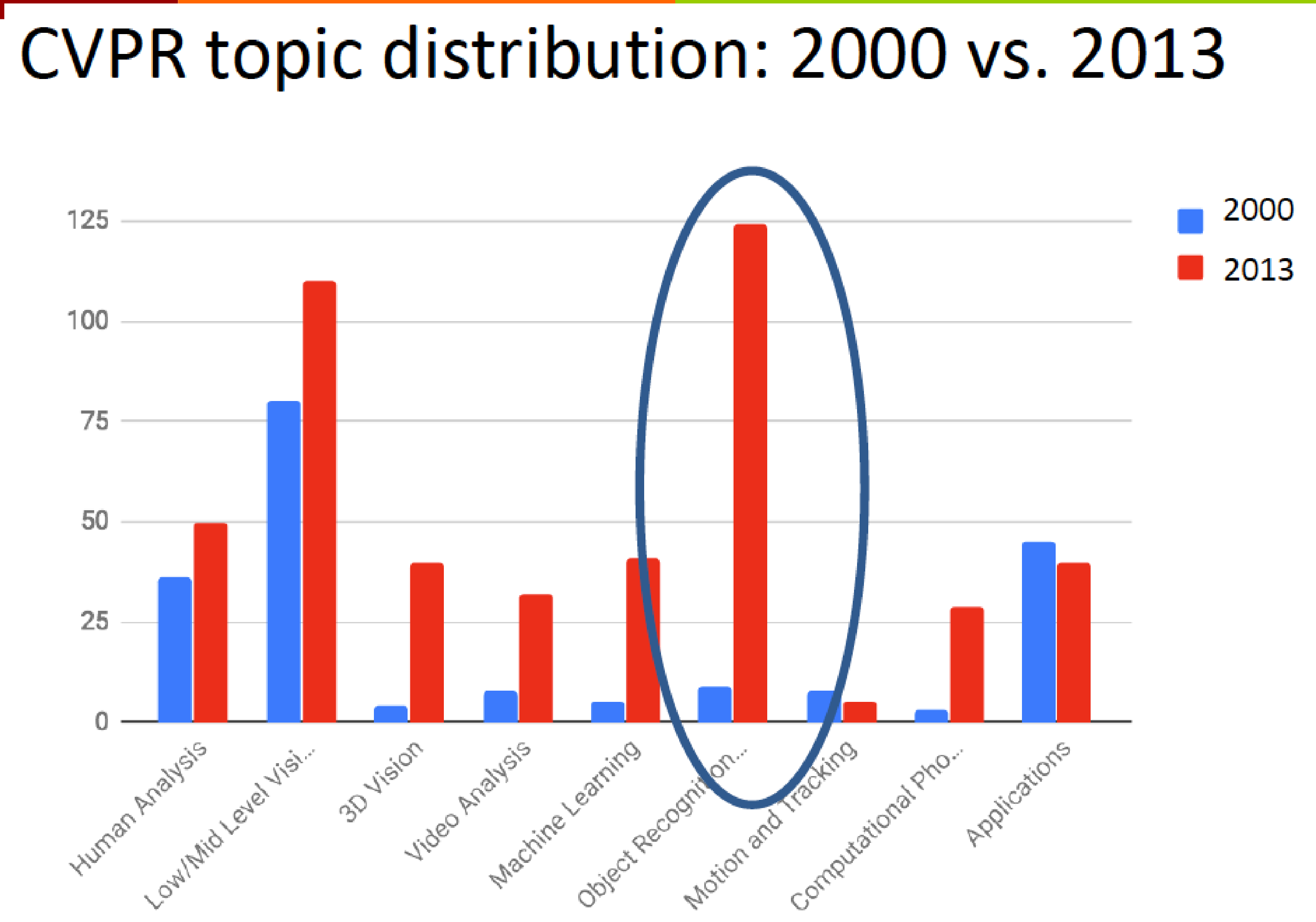
1. **"Deep Residual Learning for Image Recognition" (2016)** *Proceedings of the IEEE/C* *Computer Vision and Pattern Recognition* 25,256 citations
2. **"Deep learning" (2015)** *Nature* 16,750 citations
3. **"Going Deeper with Convolutions" (2015)** *Proceedings of the IEEE/CVF Conference on Computer 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 topic distribution: 2000 vs 2013



CVPR'2022

MENU ▾ Select Primary Subject Area

	Select Primary Subject Area	# papers	Oral	Poster
1	Recognition: detection, categorization, retrieval	177	25	152
2	Image and video synthesis and generation	157	26	131
3	3D from multi-view and sensors	137	26	111
4	Low-level vision	110	19	91
5	Vision + language	105	20	85
6	Segmentation, grouping and shape analysis	99	16	83
7	Transfer/ low-shot/ long-tail learning	86	15	71
8	Deep learning architectures and techniques	85	20	65
9	Self-& semi-& meta- & unsupervised learning	84	7	77
10	Video analysis and understanding	77	15	62
11	Pose estimation and tracking	62	14	48
12	Representation learning	61	11	50
13	3D from single images	60	10	50
14	Scene analysis and understanding	56	9	47
15	Face and gestures	54	7	47
16	Computational photography	53	10	43
17	Motion and tracking	53	8	45
18	Adversarial attack and defense	52	10	42
19	Datasets and evaluation	52	7	45
20	Machine learning	41	7	34
21	Action and event recognition	40	8	32
22	Efficient learning and inferences	40	3	37
23	Medical, biological and cell microscopy	37	5	32
24	Vision applications and systems	32	4	28
25	Navigation and autonomous driving	31	2	29
26	Vision + graphics	24	6	18
27	Privacy and federated learning	21	3	18
28	Vision + X	20	4	16
29	Physics-based vision and shape-from-X	16	2	14
30	Robot vision	16	3	
31	Explainable computer vision	15	2	13
32	Demo	15		15
33	Optimization methods	14	2	12
34	Transparency, fairness, accountability, privacy and ..	14	4	
35	Document analysis and understanding	12		12
36	Biometrics	11	2	9

Categories at CVPR 2022 ranked by number of papers accepted

01:42

CVPR'2022


transformer
video
based
object
image
detection
segmentation
aware
3d
semantic
model
representation
recognition
supervised
self
estimation
detection object
adversarial
temporal
robust
unsupervised
efficient
visual
label
shot
monocular
prediction
contrastive
multi
language
toward
feature
domain
attention
matching
tracking
field
action
vision
flow
text
prior
generative
dynamic
localization
cloud
pose
synthesis
knowledge
attack
reconstruction
scene
global
architecture
face
hierarchical
semi supervised
implicit
correspondence
style
motion
understanding
dataset
depth
classification
view
local
transfer
human
distillation
task
training
dense
shape
framework
end
generation
point
adaptive
distribution
sparse
stereo guided
graph
data
camera
alignment
large scale
weakly
cross

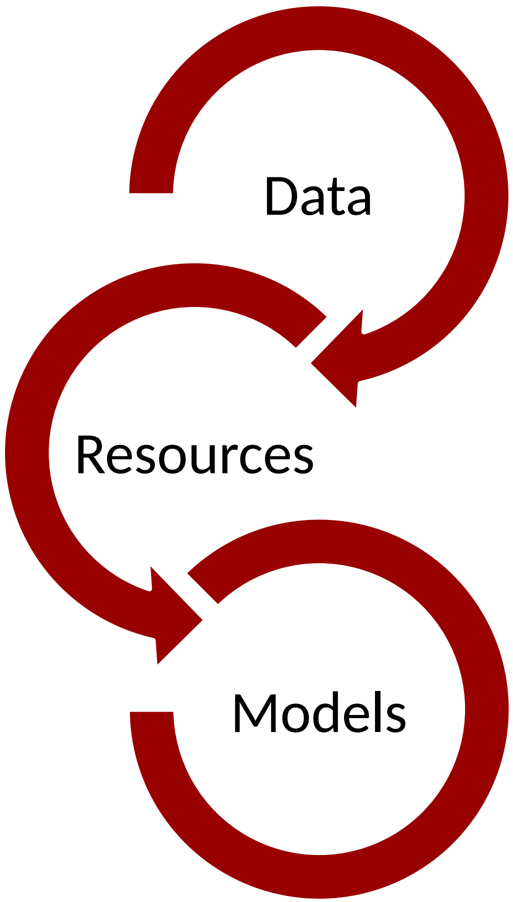
Deep learning – where does it come from?

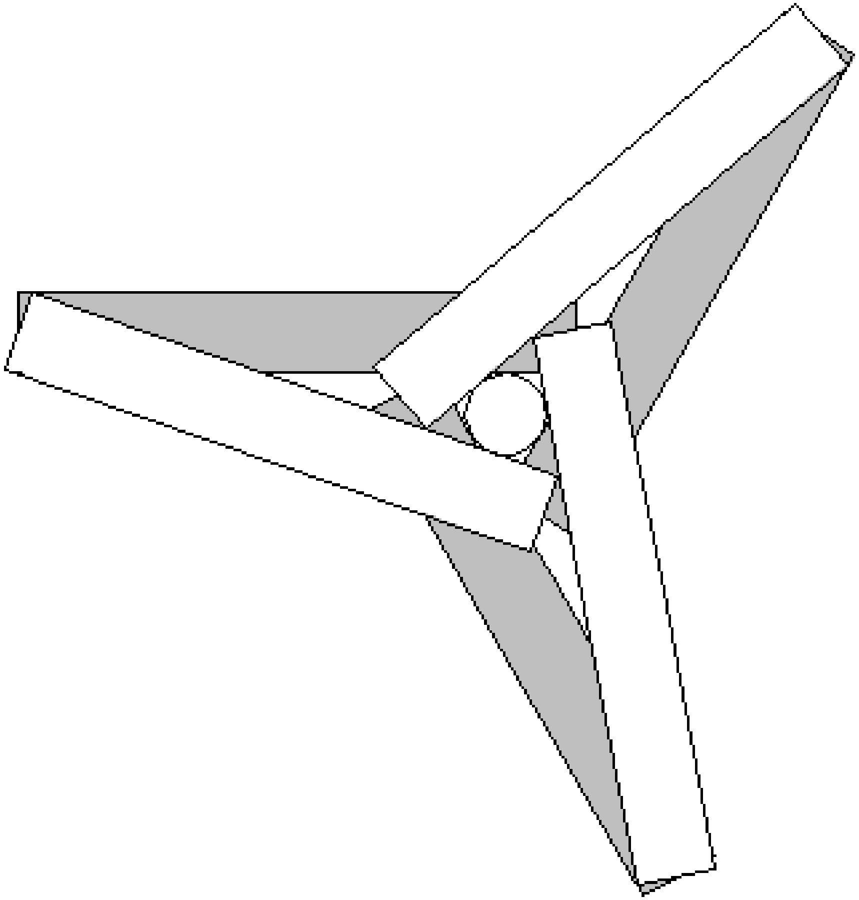
- **1943:** neurophysiologist Warren McCulloch and mathematician Walter Pitts – a **neuronal model** using an electrical circuit
- **1950**: Alan Turing created the world-famous **Turing Test**: a computer to pass, has to be able to convince a human that it is a human and not a computer.
- **1952**: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

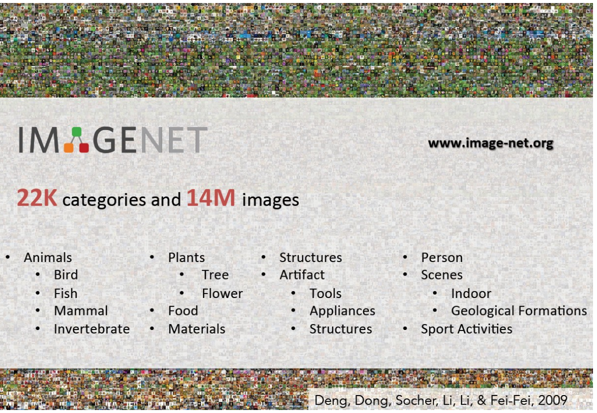
90% of all digital data were generated last 2 years.

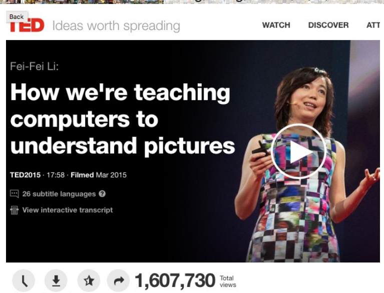
Daily:

- 300M photos get uploaded
- <u>95M photos and videos</u> are shared on Instagram
- <u>100M people</u> use the Instagram “stories”
- 15K GIFs are sent via Facebook
- 154K calls on Skype
- 4.7T photos stored in cameras

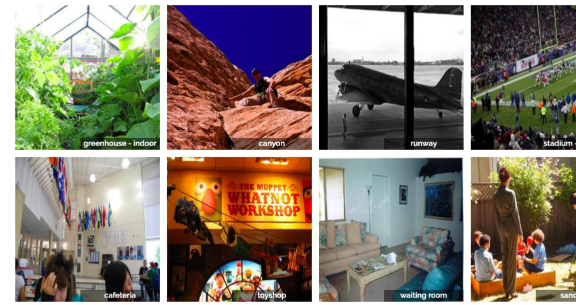
https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#46be238160ba

DL Datasets

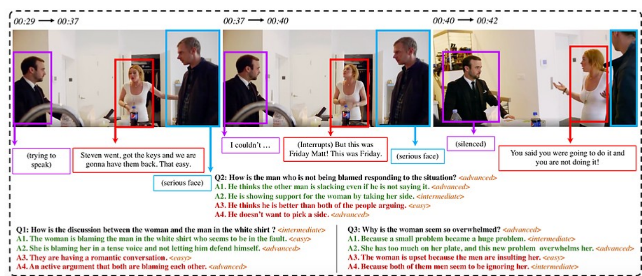




LVIS Challenge:
2.2M masks, 16K images



Places2: 10M images



SocialIQ

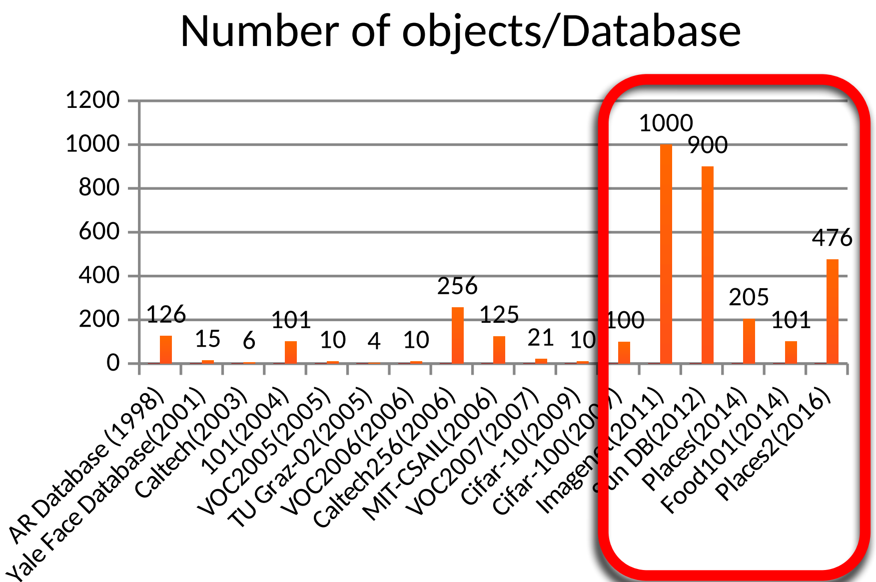
TACO: Waste in the wild

Lyft Level 5

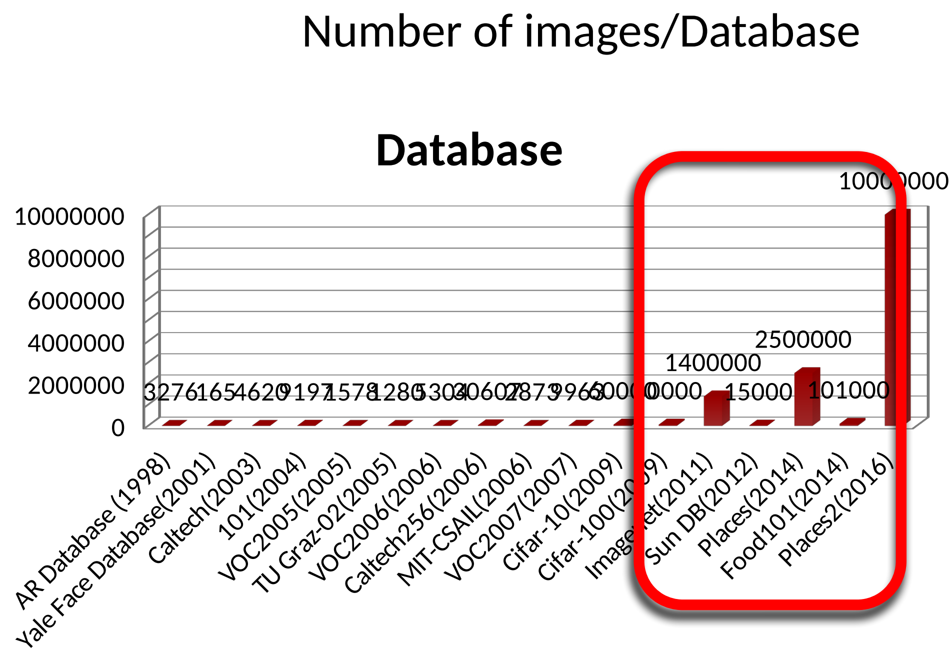
https://www.datasetlist.com

FastMRI

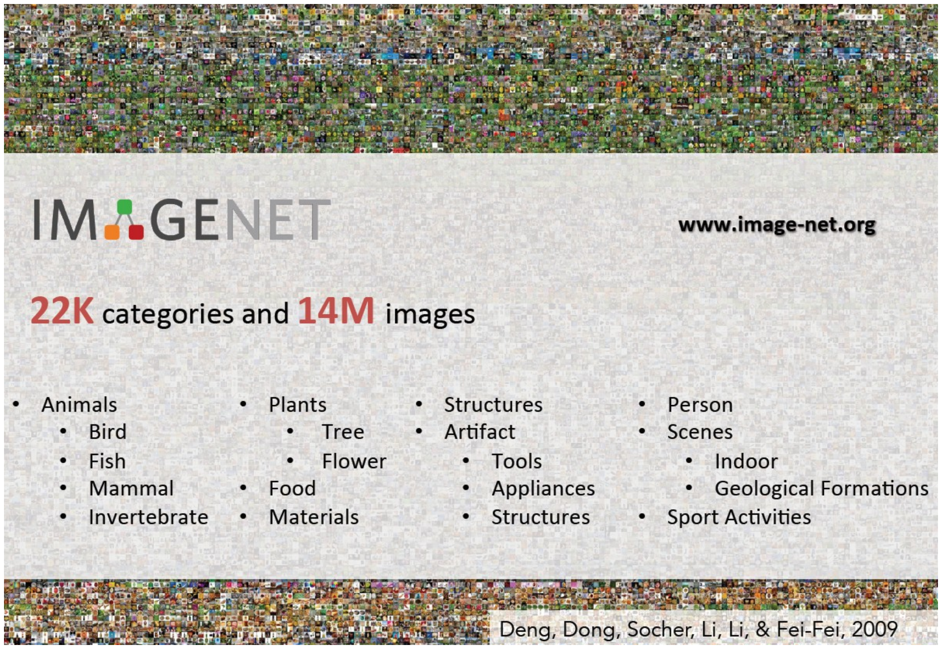
Image databases evolution

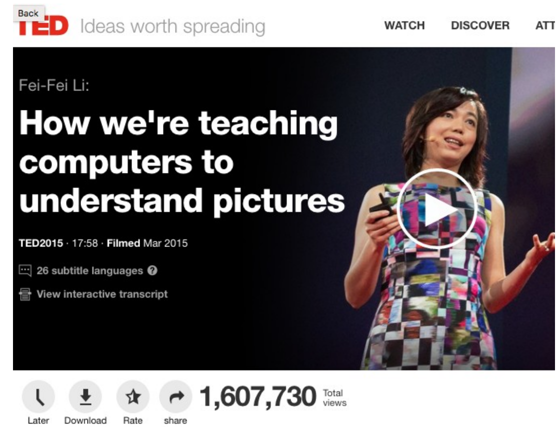


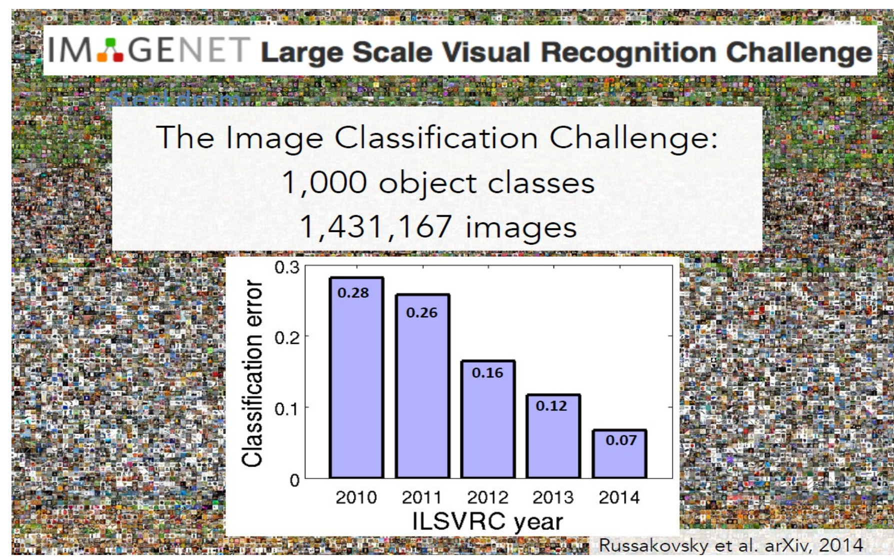
ImageNet &
Deep learning



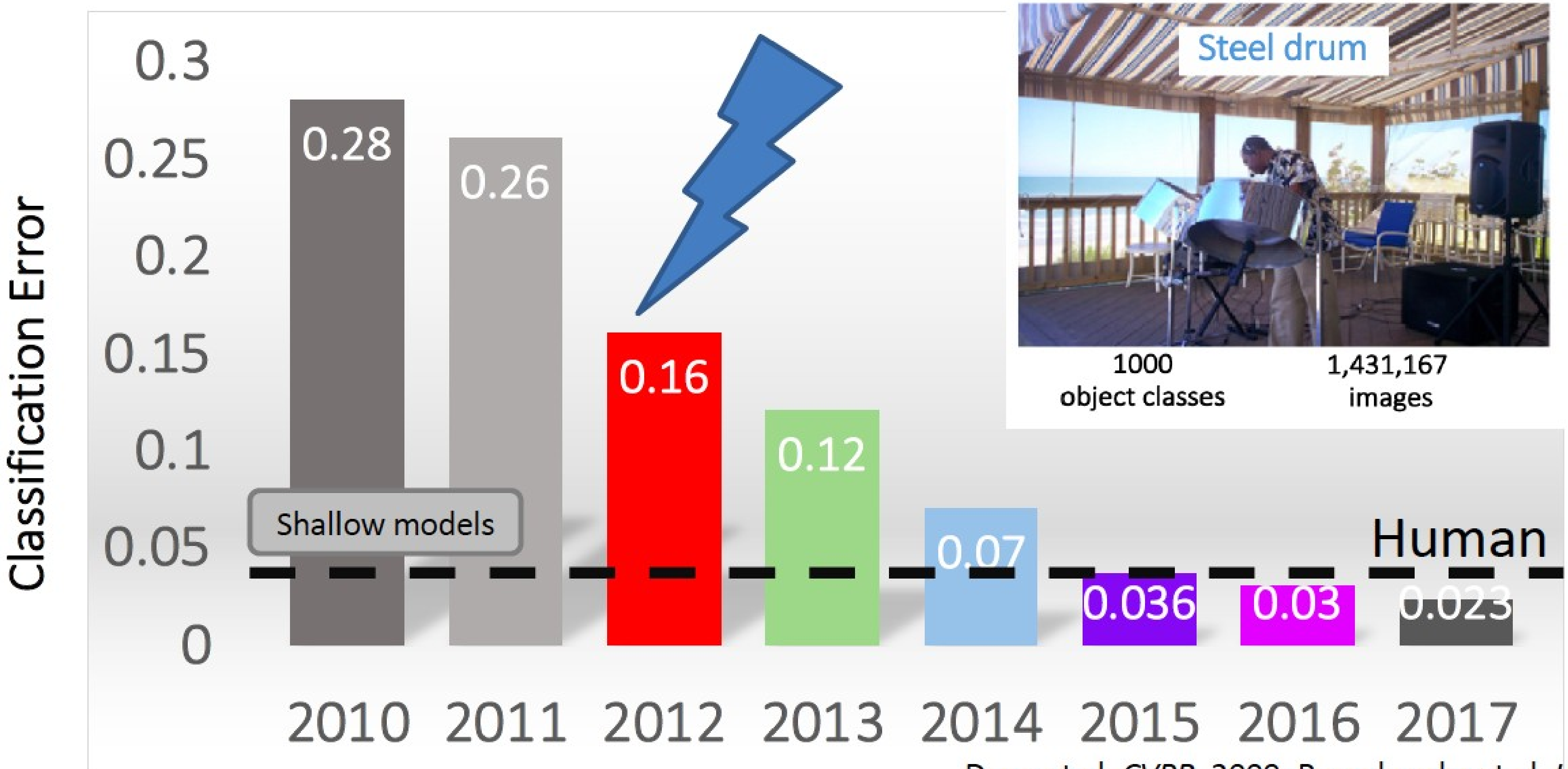
Imagenet


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


Back
TED Ideas worth spreading
WATCH
DISCOVER
ATT
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

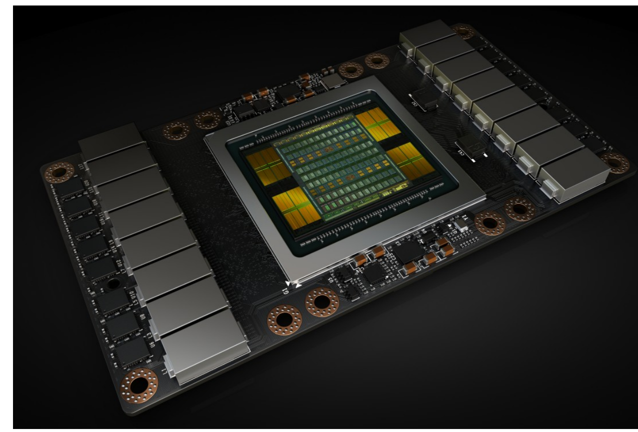
IMAGENET Classification **Task**

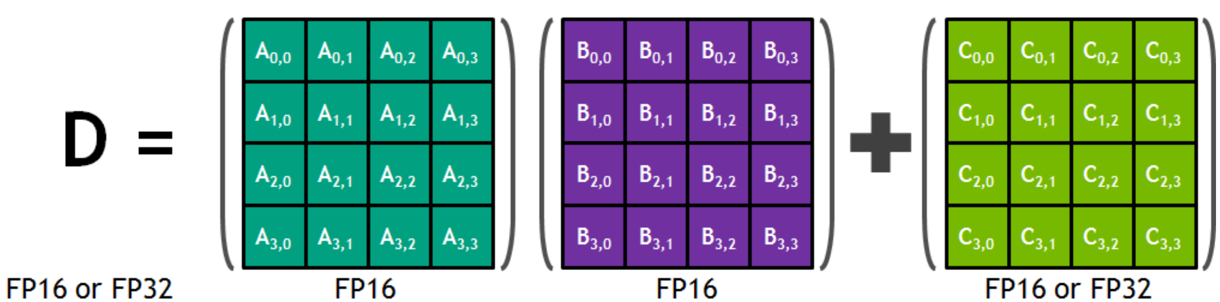


Deng et al. *CVPR*, 2009; Russakovsky et al. *IJCV*, 2012;

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)





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

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

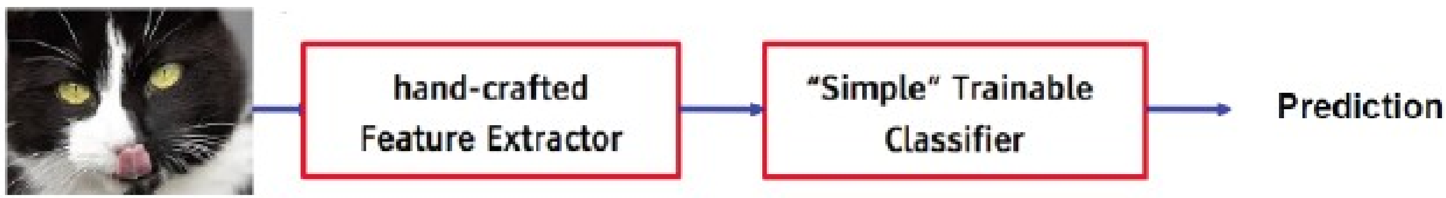
- AI, Machine learning & Deep learning

- What is a Convolutional Neural Network?
  - Layers
  - Optimization

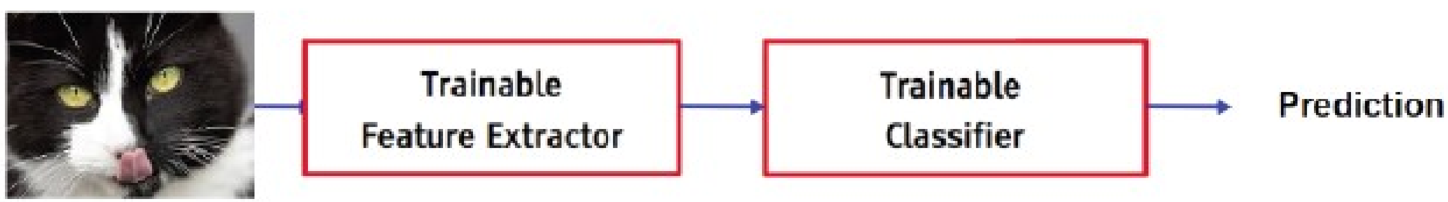
- Applications

Why Deep learning?

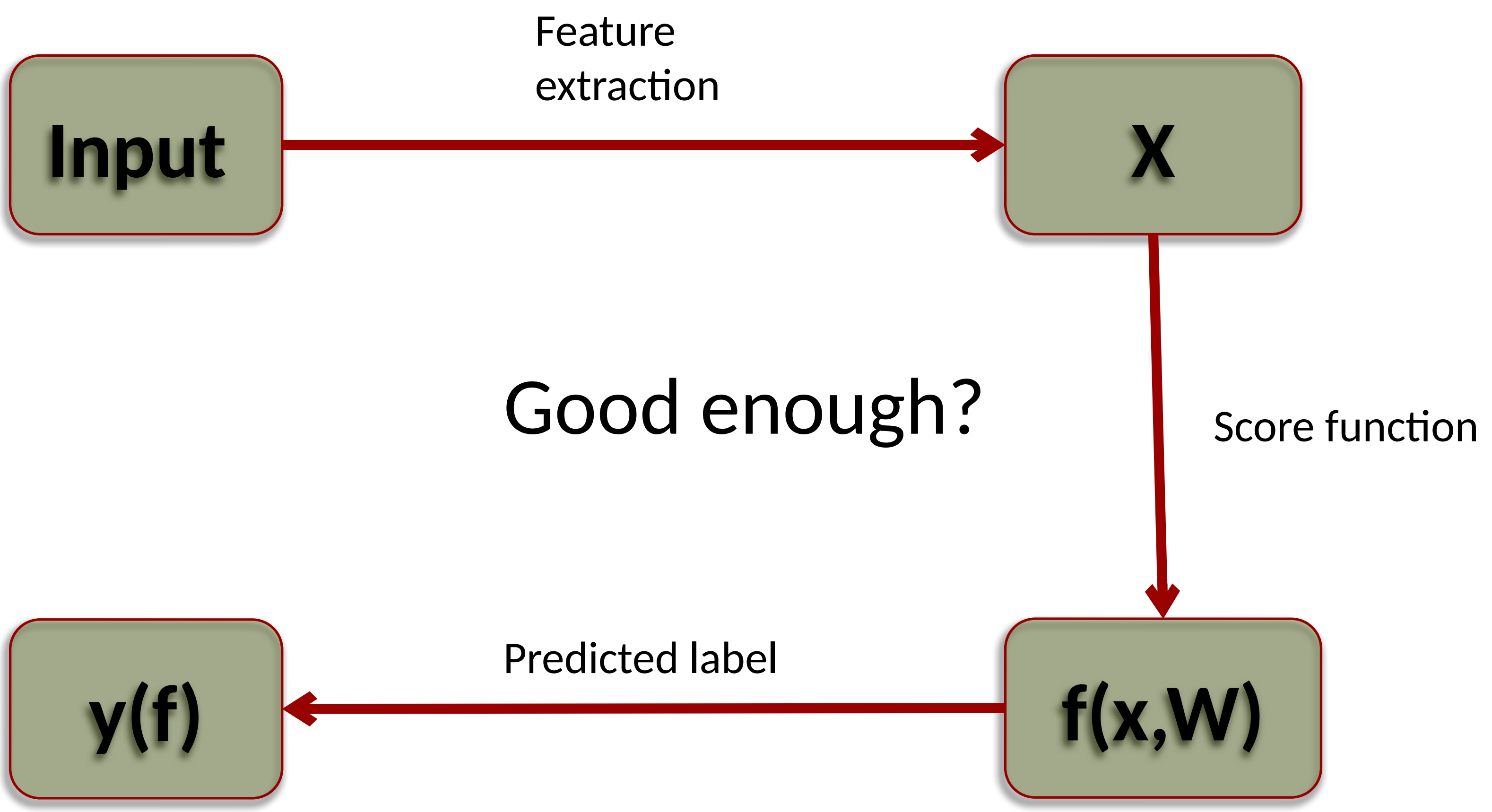
Classical way of solving Computer Vision problems



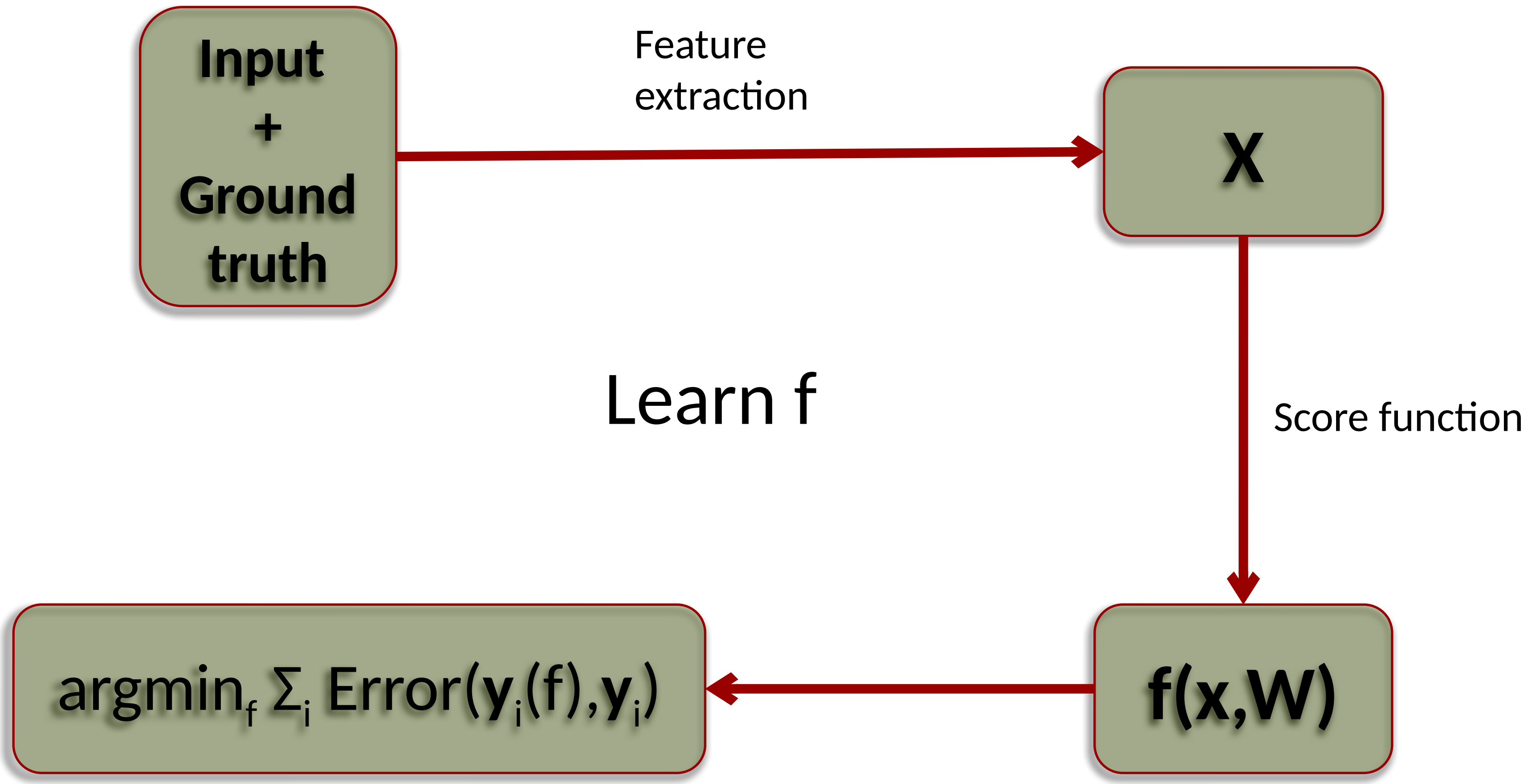
How Computer Vision problems are solved by Deep Learning



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)

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

Expectation over data distribution

Prediction

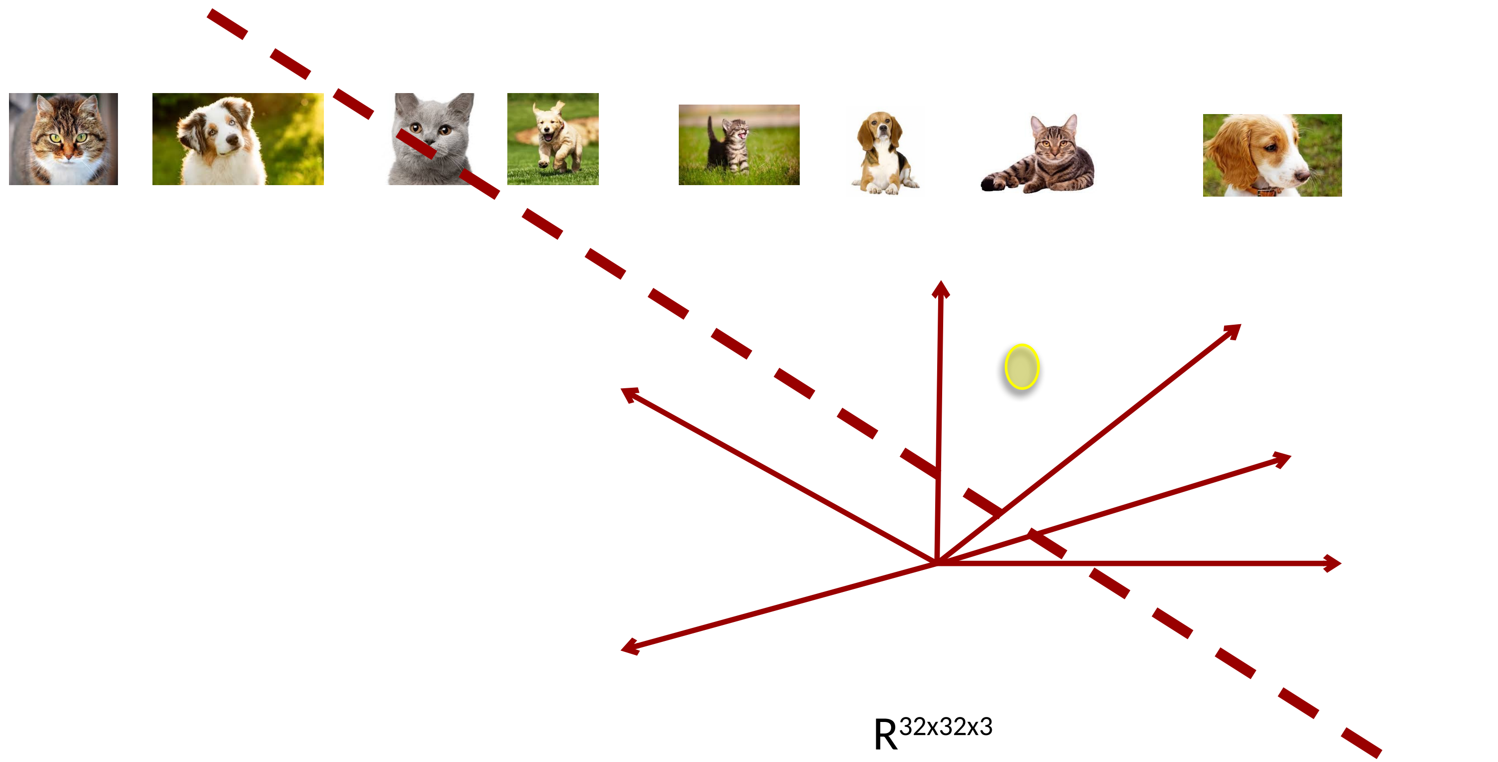
Ground Truth

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?



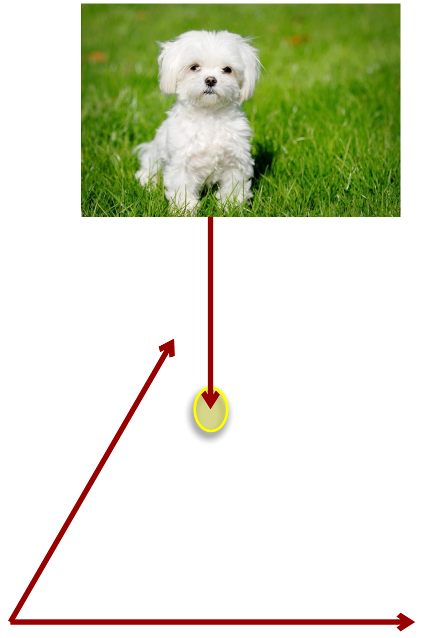
To find the hyperplane that separates dogs from cats, we need to define:

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

Linear classification

How to project data in the feature space:

$$\underbrace{f(x)}_{3\times1} = \underbrace{W}_{3\times3072}\ \underbrace{x}_{3072\times1} + \underbrace{b}_{3\times1}$$



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:

3x1 3072x1

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

3x3072 **3x1**

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

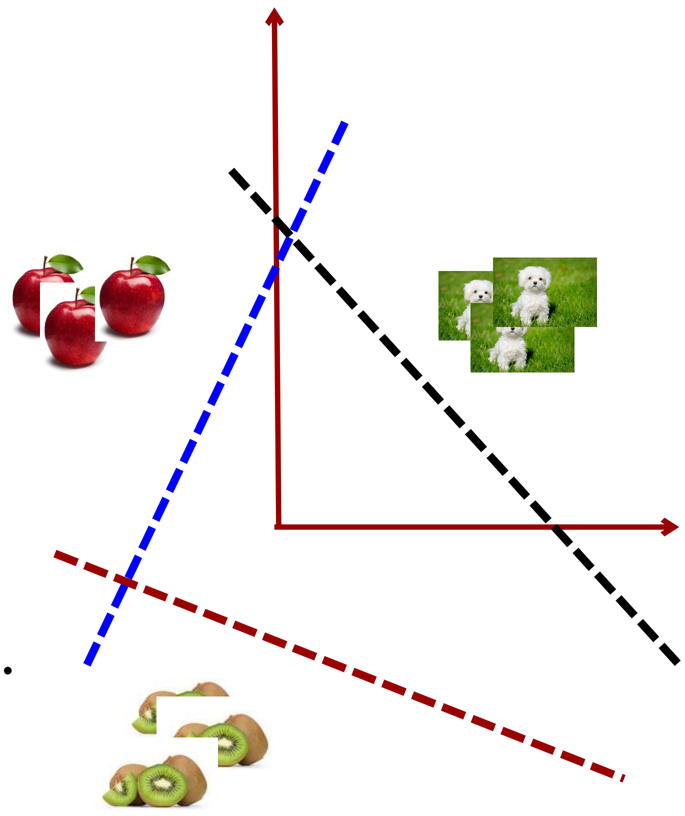
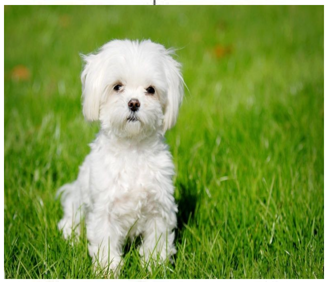


Image classification

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

stretch pixels into single column



input image

$$\underbrace{\begin{bmatrix} 0.2 & -0.5 & 0.1 & 2.0 \\ 1.5 & 1.3 & 2.1 & 0.0 \\ 0 & 0.25 & 0.2 & -0.3 \end{bmatrix}}_{W} \underbrace{\begin{bmatrix} 56 \\ 231 \\ 24 \\ 2 \end{bmatrix}}_{x_i} + \underbrace{\begin{bmatrix} 1.1 \\ 3.2 \\ -1.2 \end{bmatrix}}_{b} \longrightarrow \underbrace{\begin{bmatrix} -96.8 \\ 437.9 \\ 61.95 \end{bmatrix}}_{f(x_i; W, b)} \begin{array}{l} \text{cat score} \\ \text{dog score} \\ \text{ship score} \end{array}$$

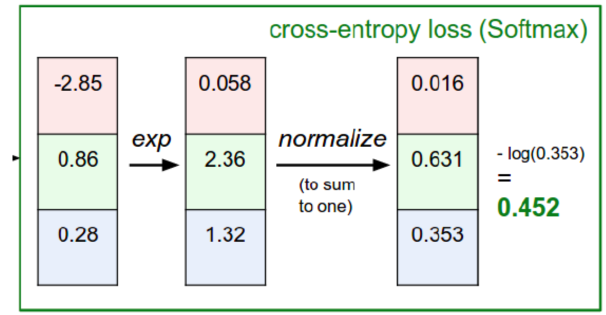
Adapted from: Fei-Fei Li & Andrej Karpathy & Justin Johnson

Loss function and optimisation

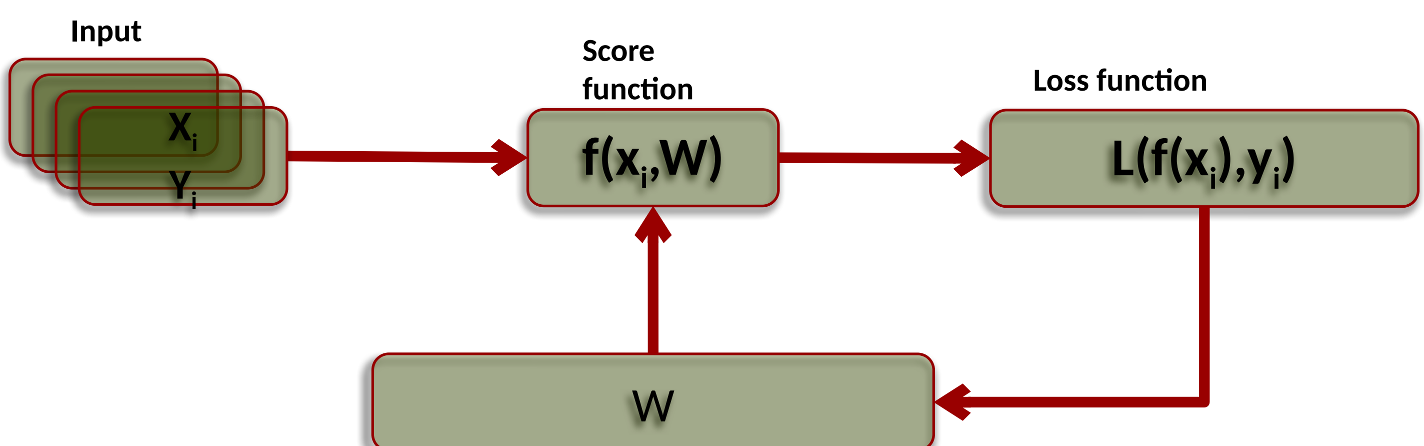
Question: if you were to assign a single number to how unhappy you are with these scores, what would you do?

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

softmax function



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

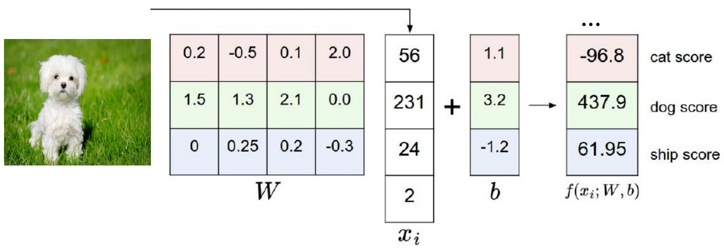


Index

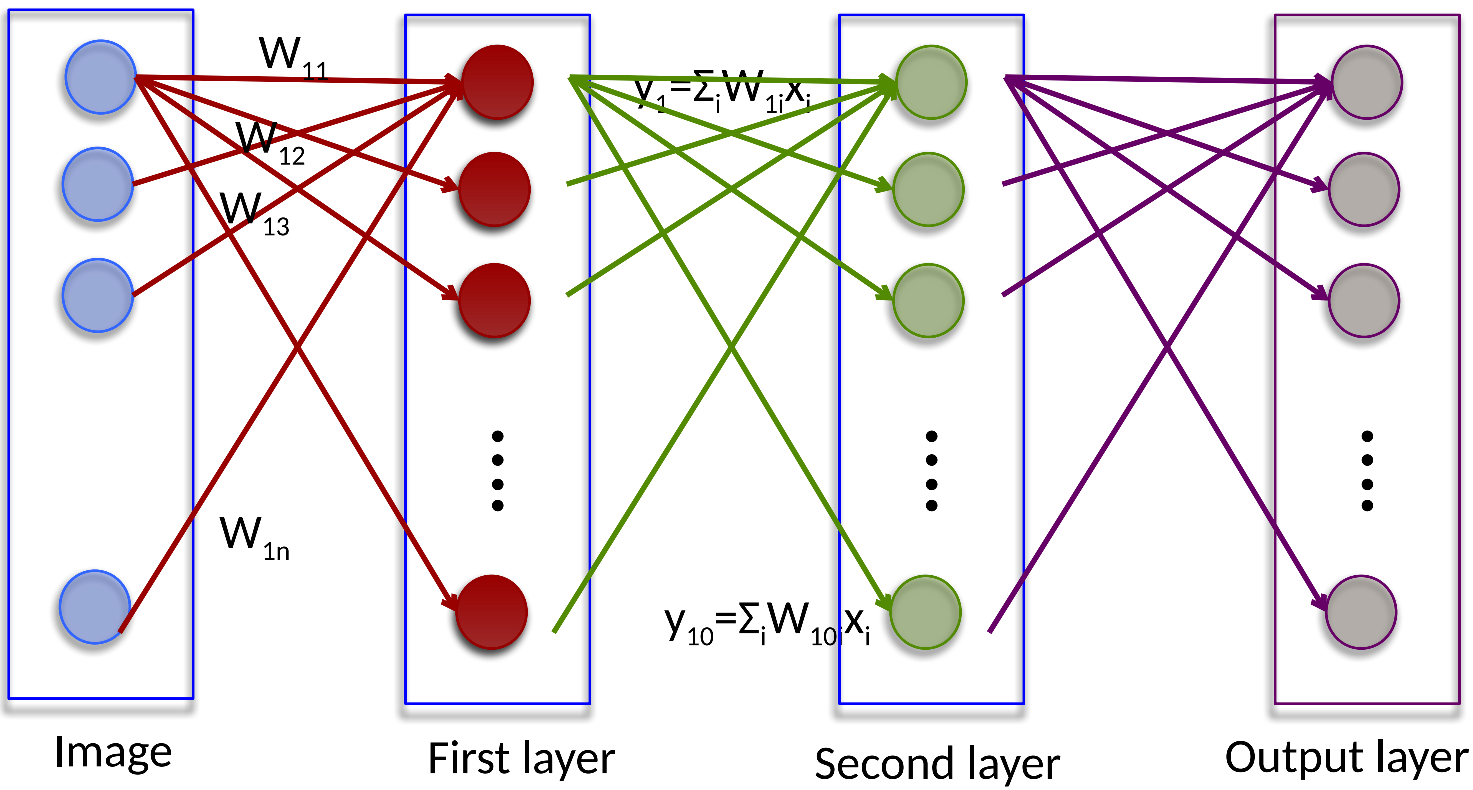
- AI, Machine learning & Deep learning
- What is a Convolutional Neural Network?
  - Layers
  - Optimization
- Applications

How is a CNN doing deep learning?

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

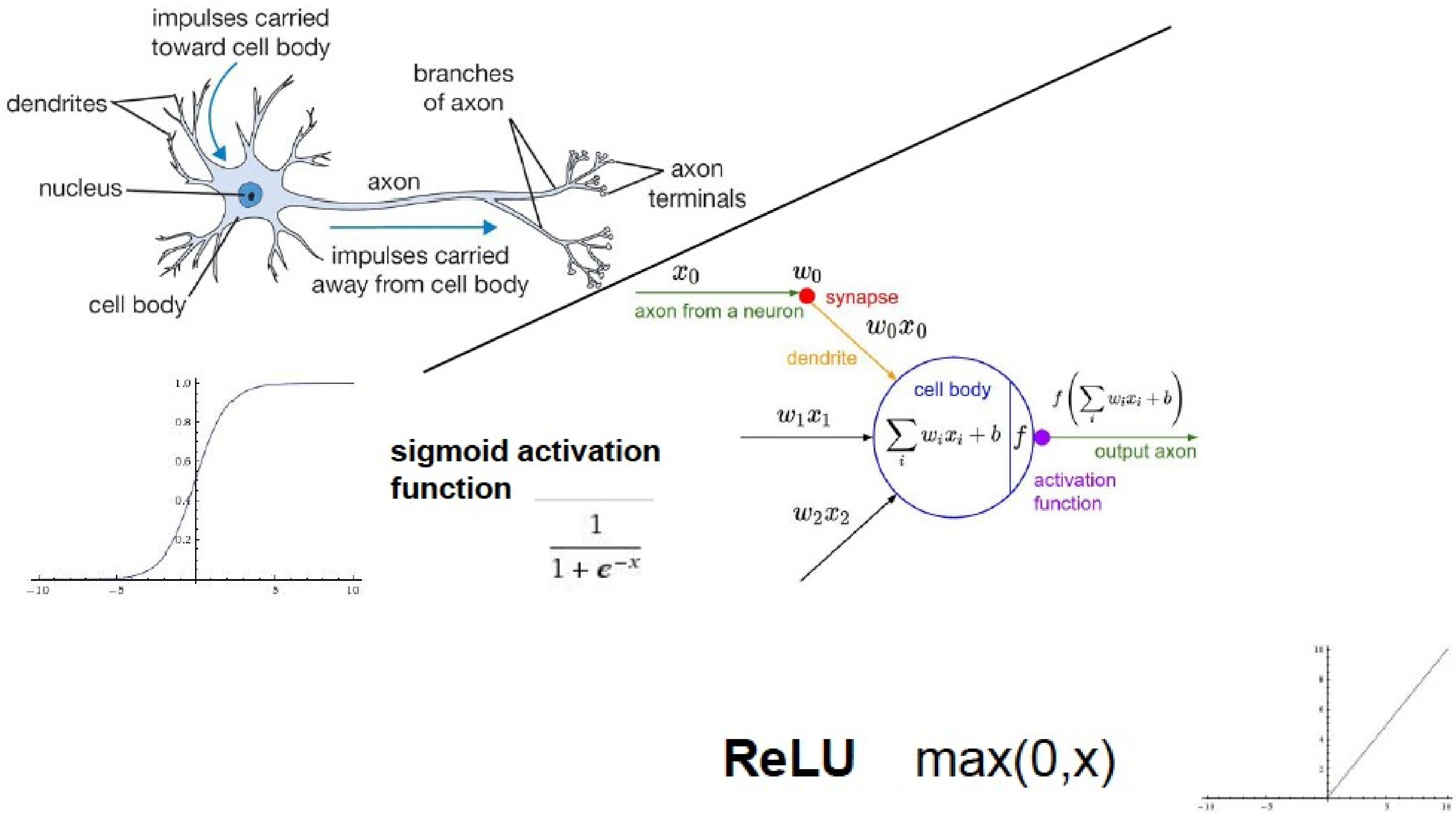


Fully connected layers



01:42

Why a CNN is a neural network?

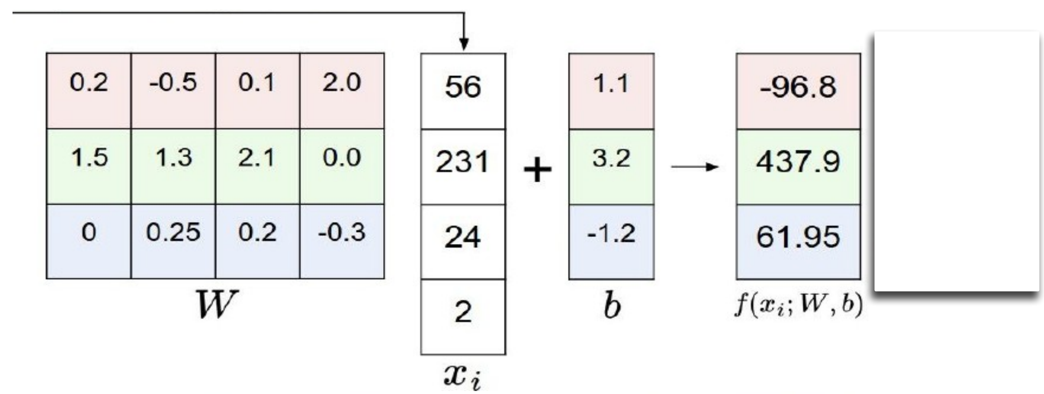


Modern CNNs – 10M neurons
Human CNNs – 5B of neurons.

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

Why is it convolutional?

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

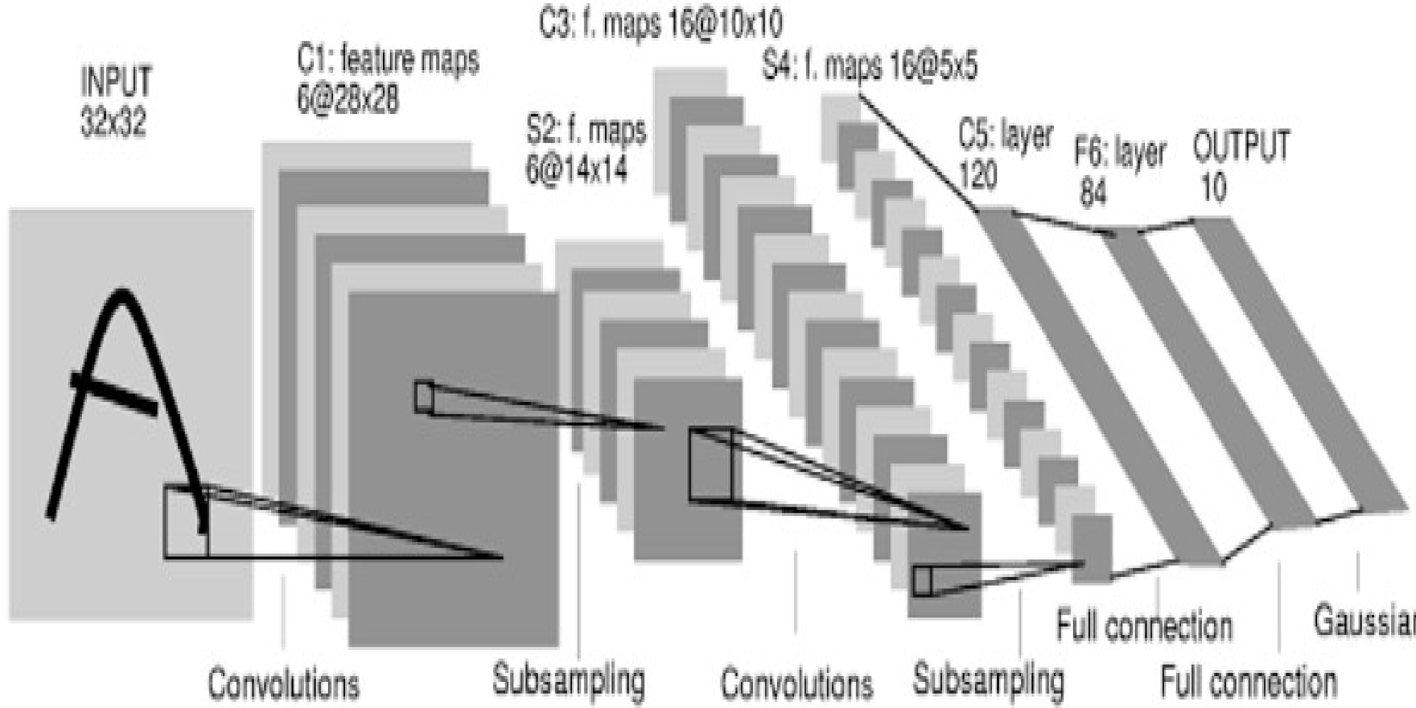


Adapted from: Fei-Fei Li & Andrej Karpathy & Justin Johnson

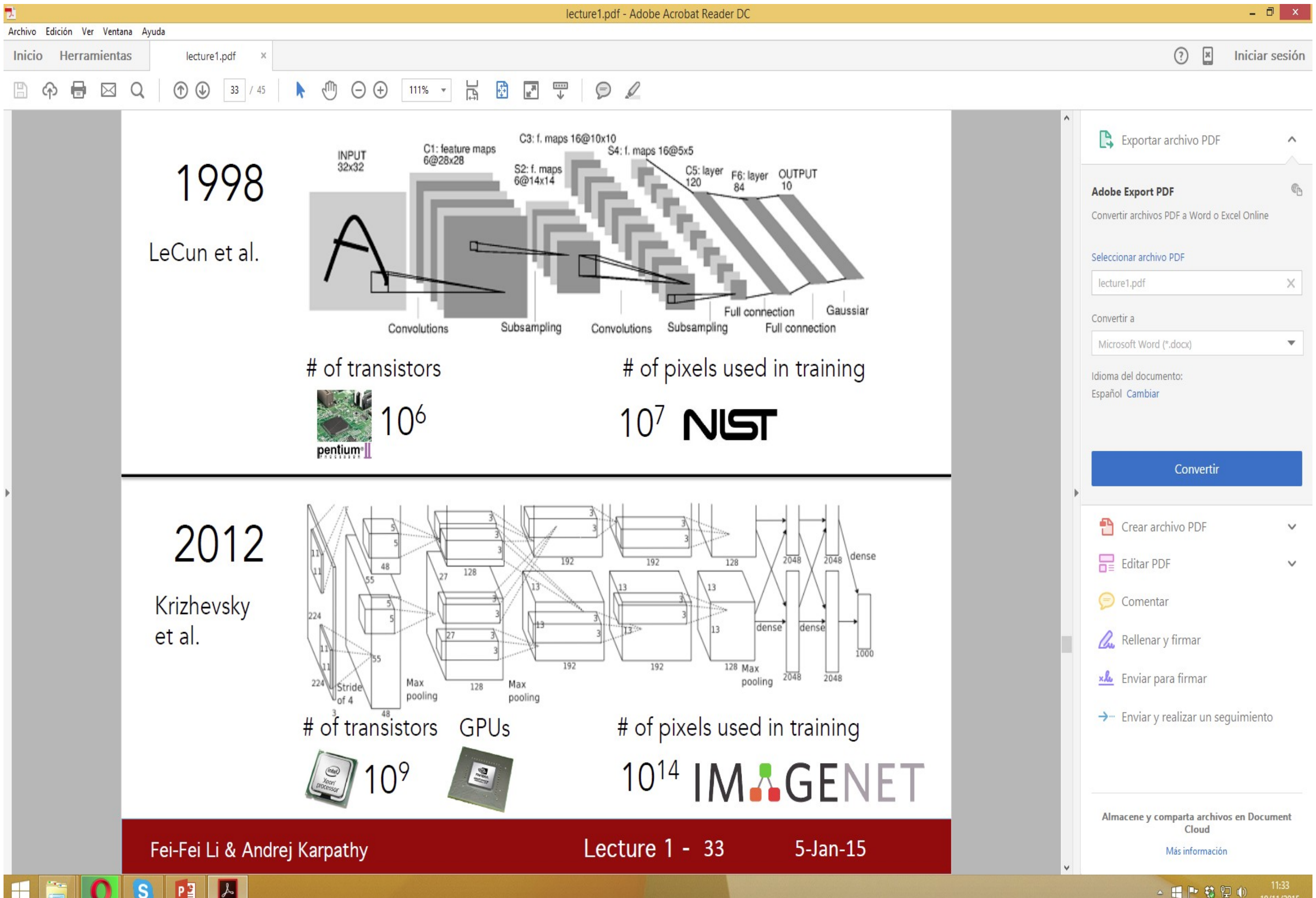
What is new in the Convolutional Neural Network?

1998

LeCun et al.



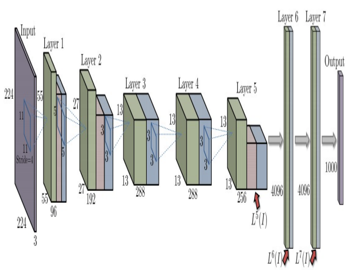
CNN evolution

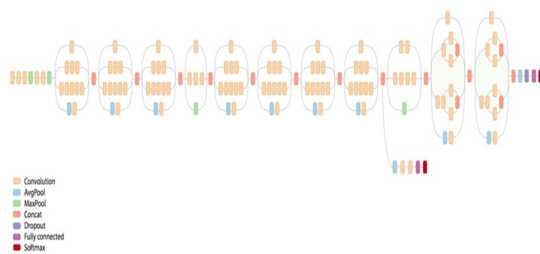


01:42

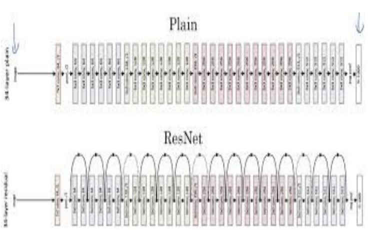
NN models

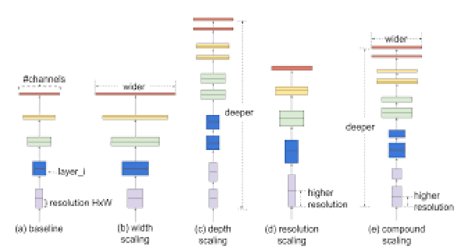
))))

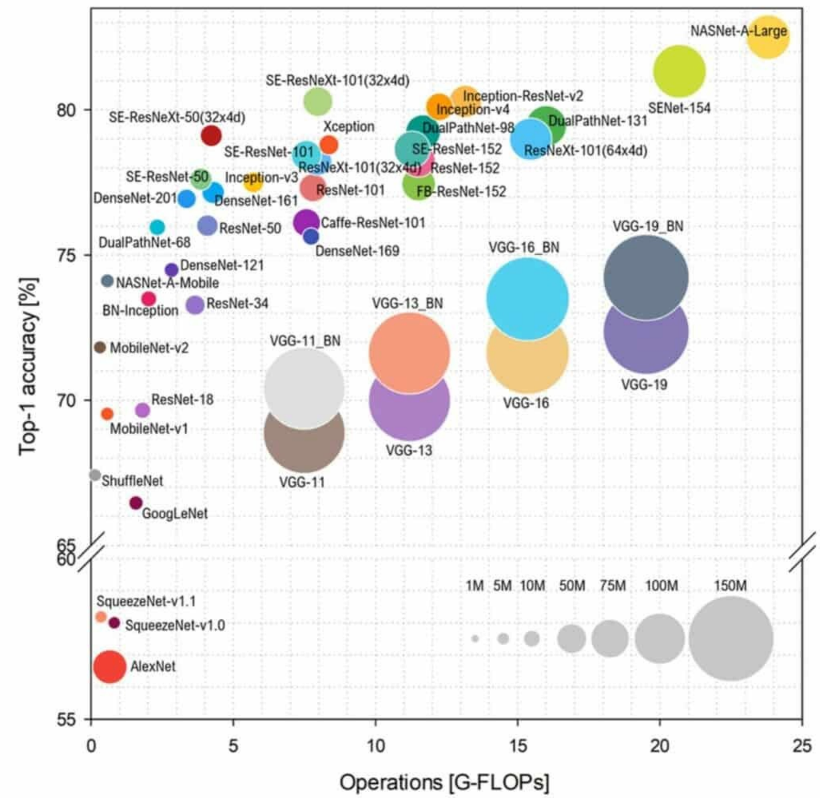




Another view of GoogLeNet's architecture.



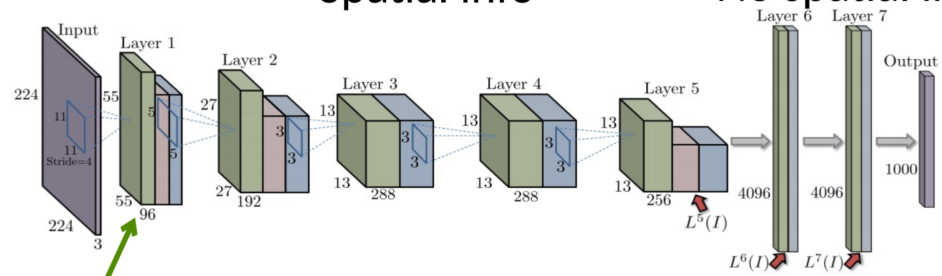


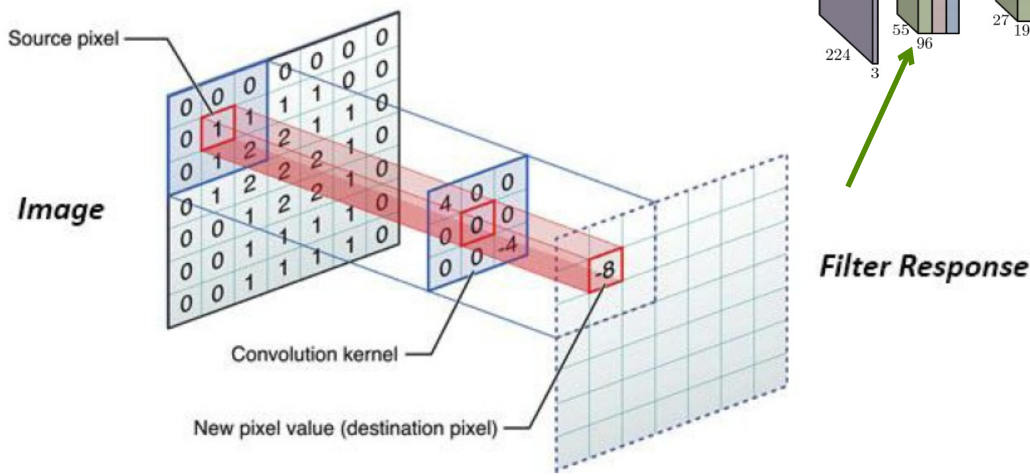


Convolutional and Max-pooling layer

Spatial info

No spatial info

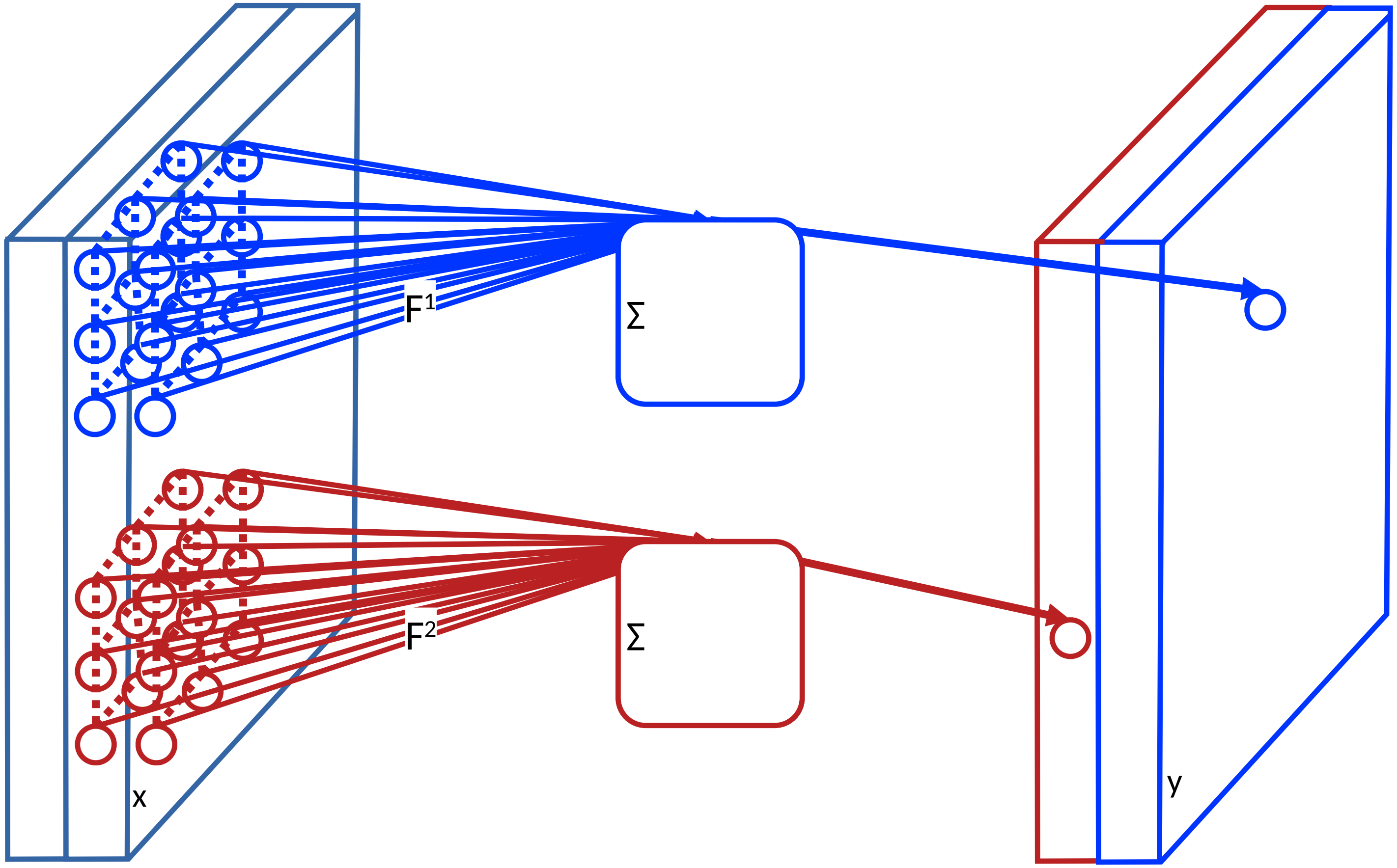




Convolutional layer

Why is it convolutional?

input features	a bank of 2 filters	2-dimensional output features

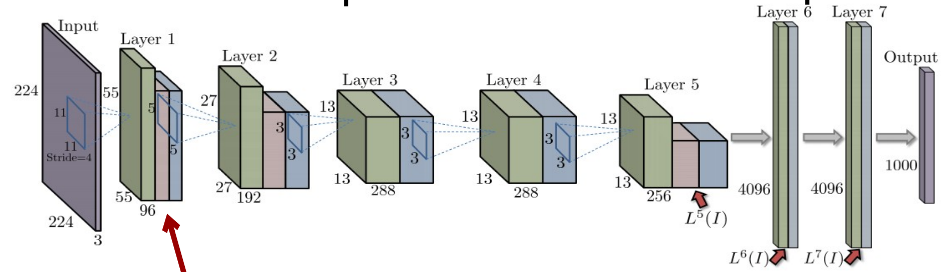


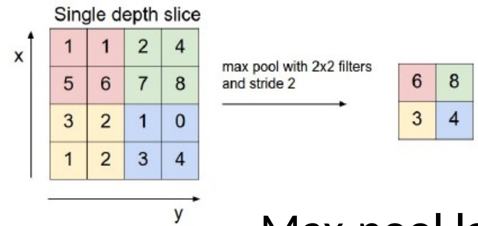
Convolution

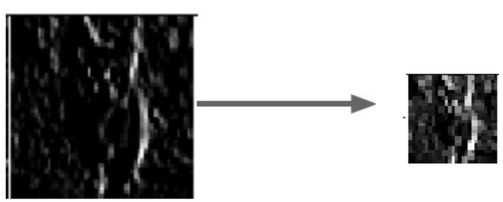
Convolutional and Max-pooling layer

Spatial info

No spatial info





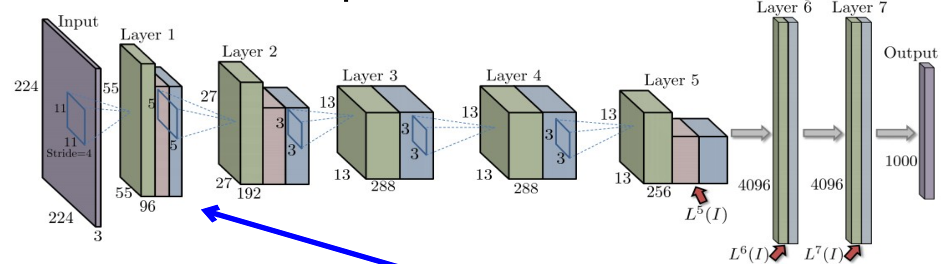


Max-pool layer

Convolutional and Max-pooling layer

Spatial info

No spatial info

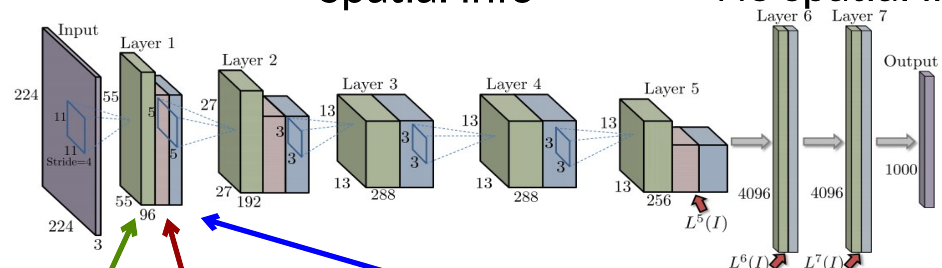


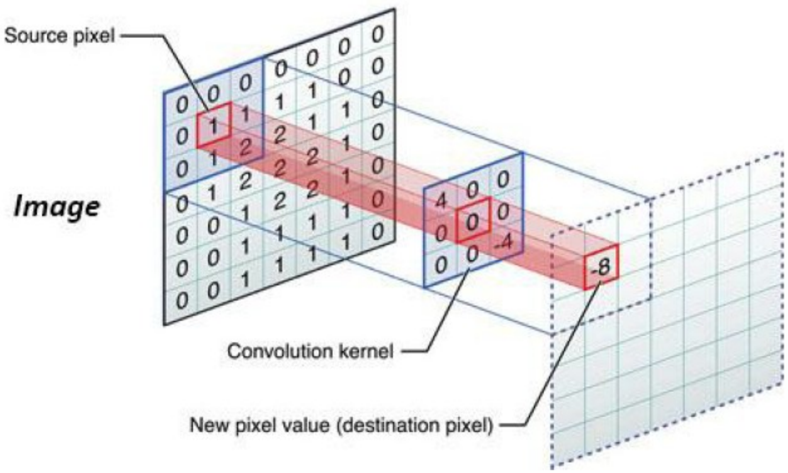
ReLU max(0,x)

Convolutional and Max-pooling layer

Spatial info

No spatial info

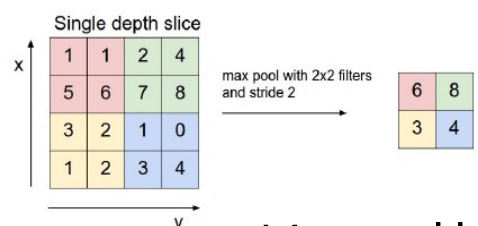




Filter Response

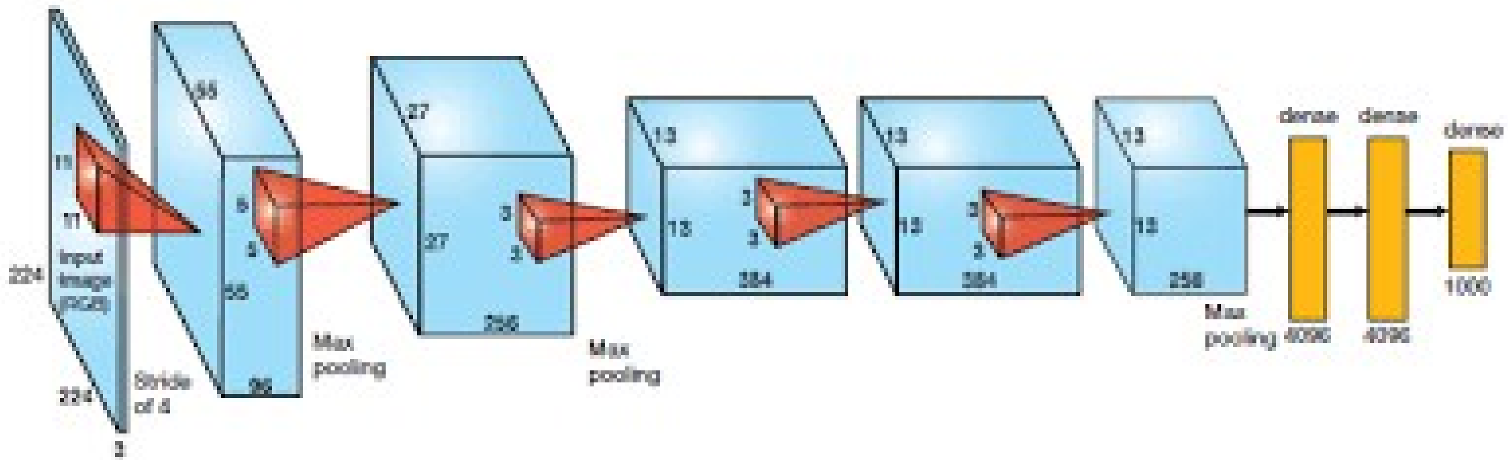
ReLU max(0,x)

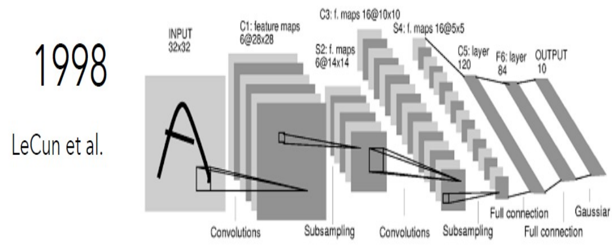
Convolutional layer



Max-pool layer

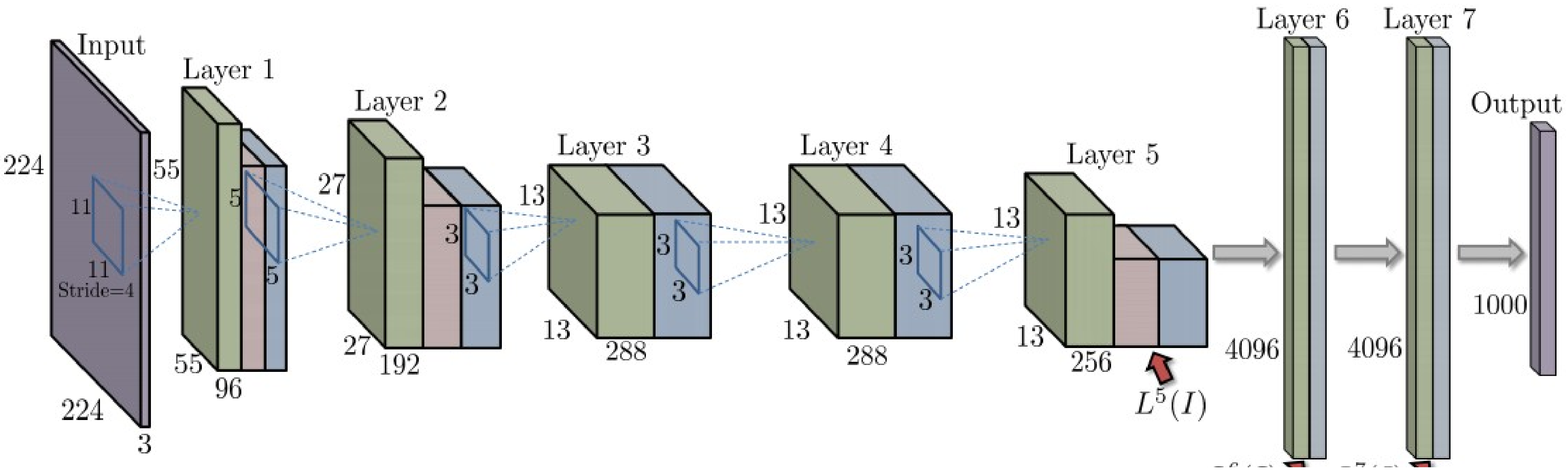
Why is it convolutional?





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

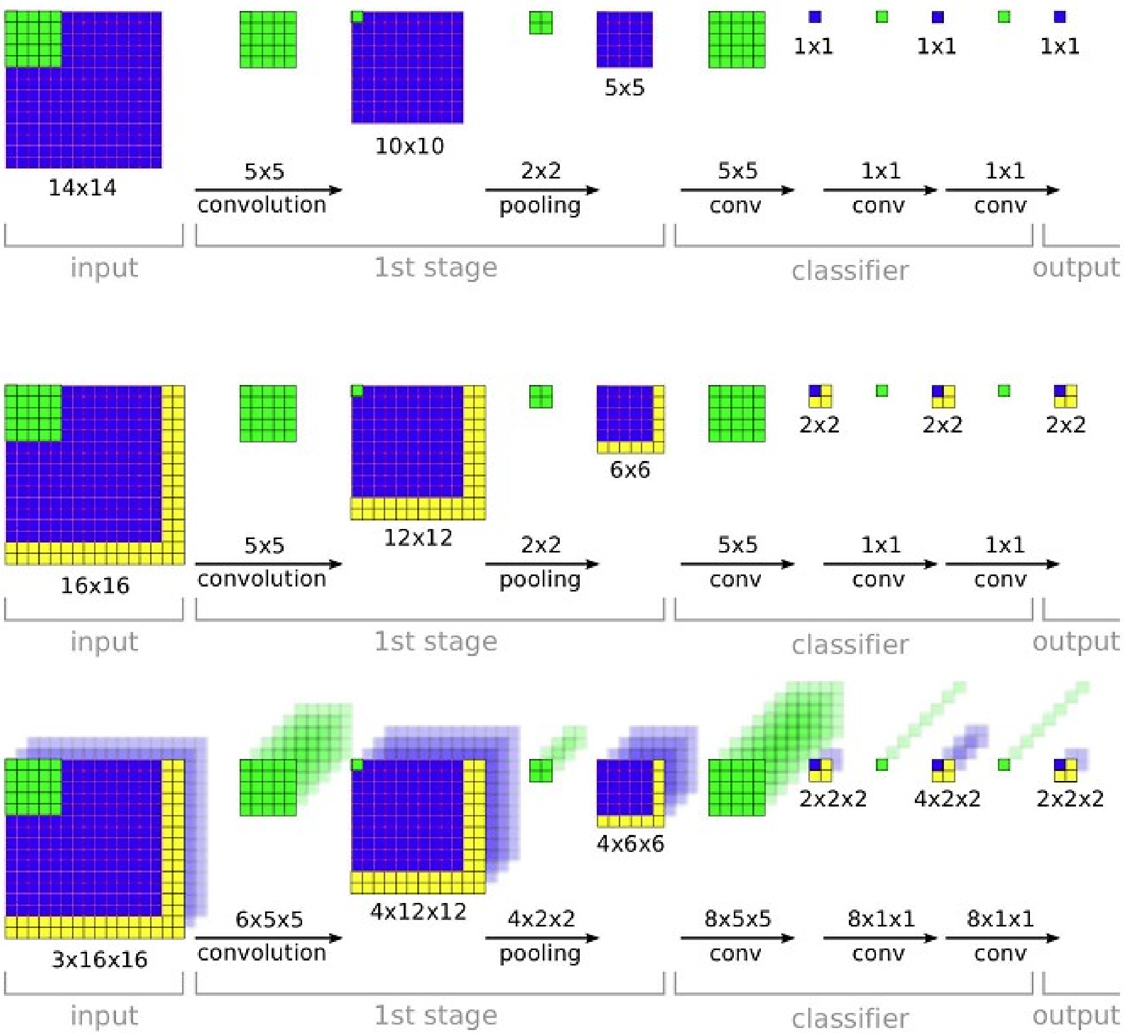
Training a CNN



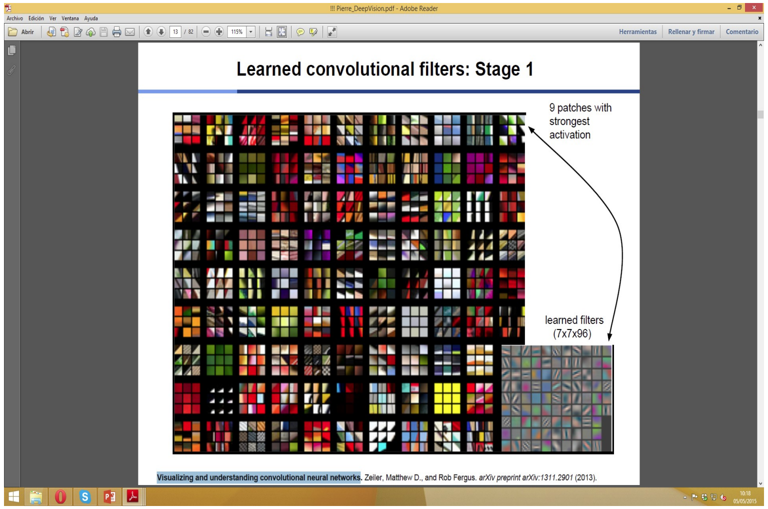
The process of training a CNN consists of training all hyperparameters: convolutional matrices and weights of the fully connected layers.

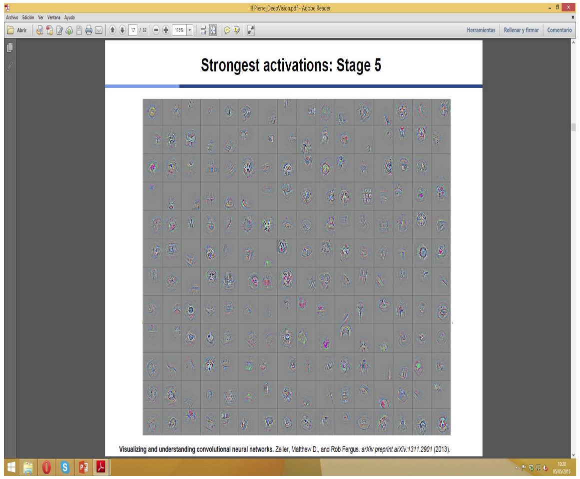
- **Several millions of parameters!!!**

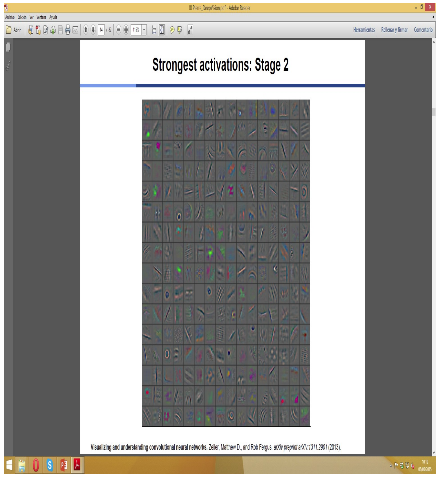
How does the CNN work?



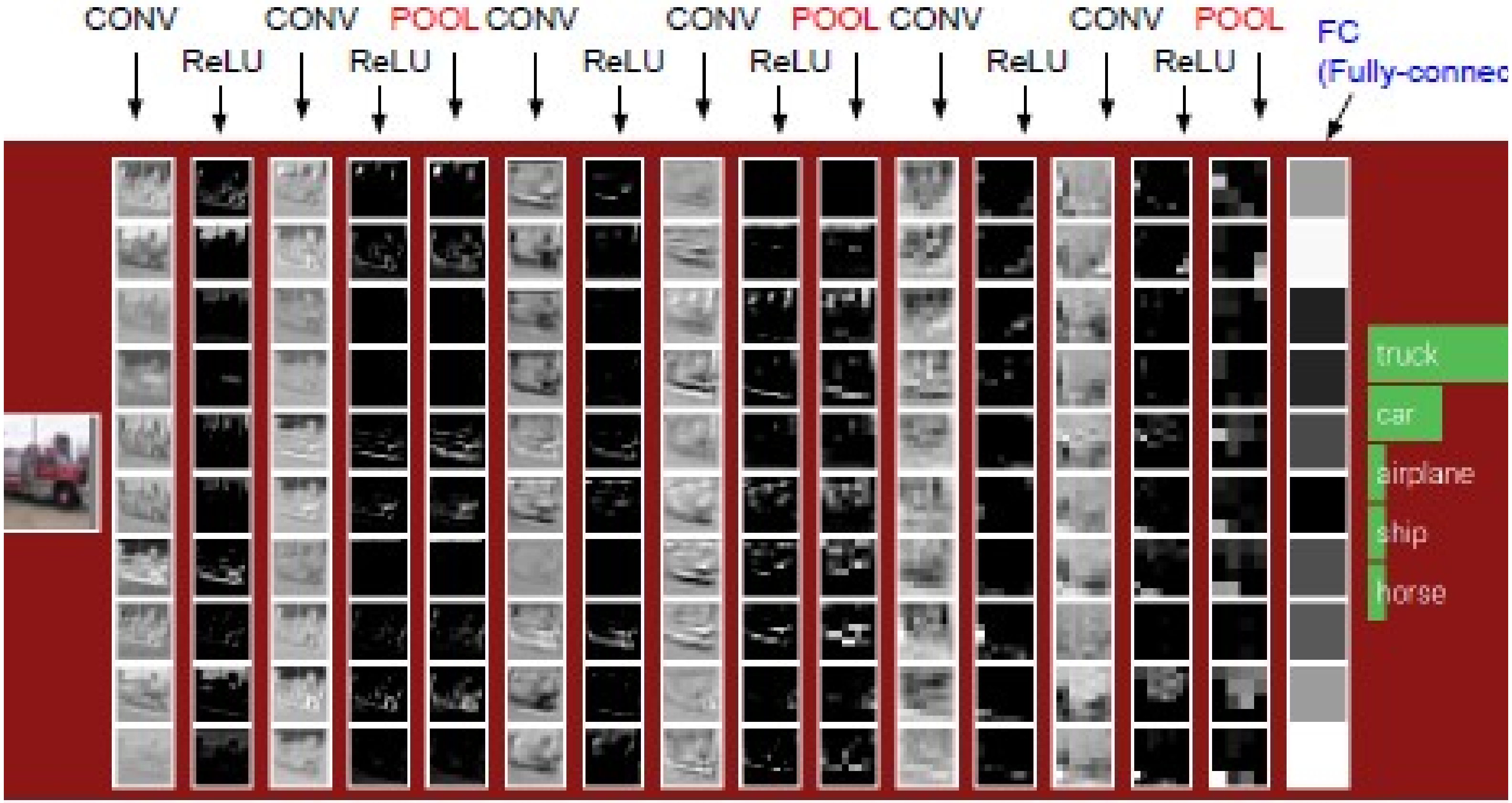
Learned convolutional filters







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 I

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

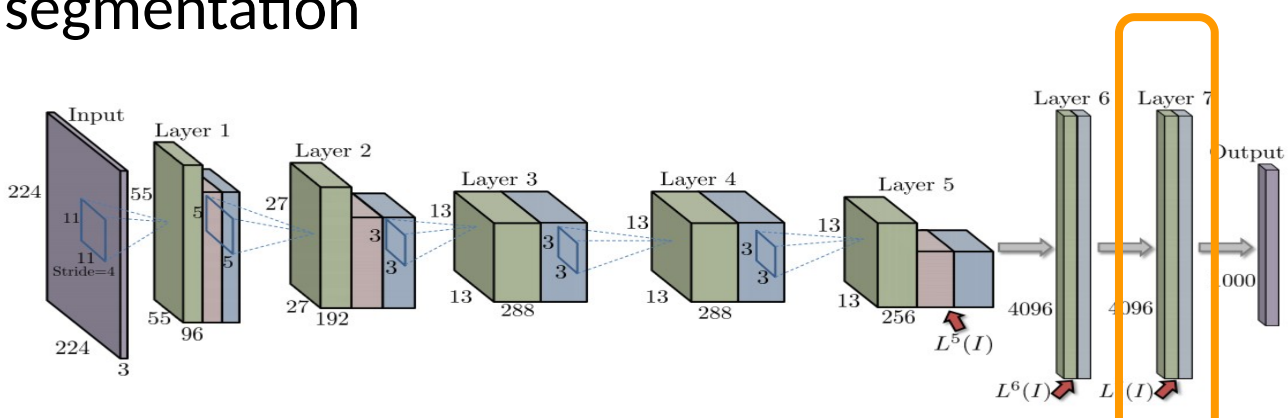
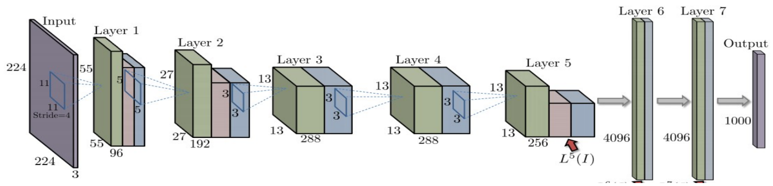


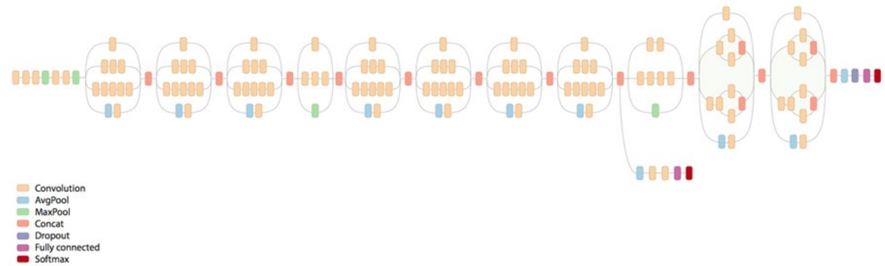
Image similarity and matching by CNN

Convolutional Neural Networks (4096 Features)

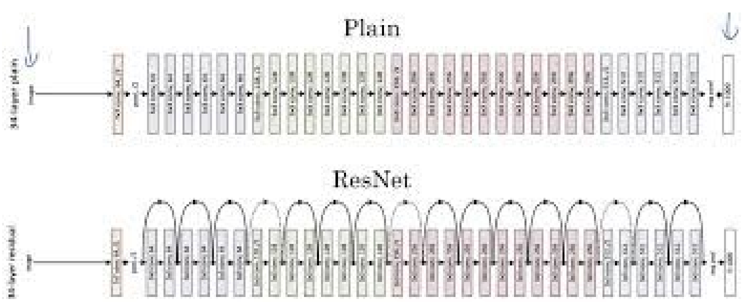
CNN models

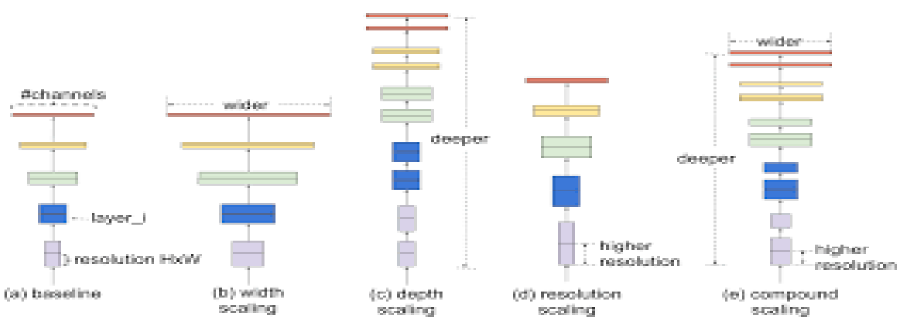
))))





Another view of GoogLeNet's architecture.

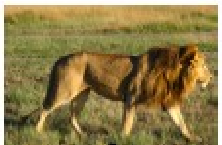


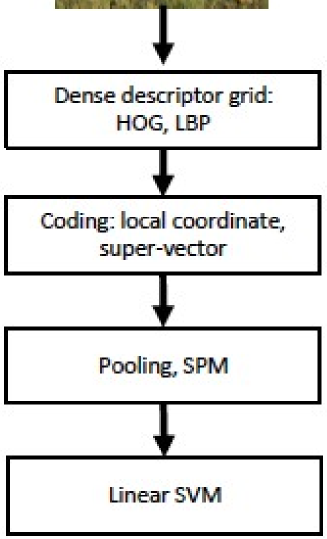


IMAGENET Large Scale Visual Recognition Challenge

Year 2010

NEC-UIUC

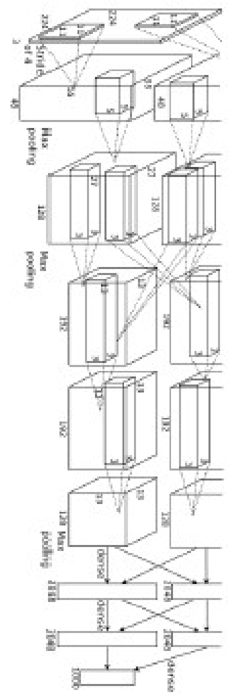




[Lin CVPR 2011]

Year 2012

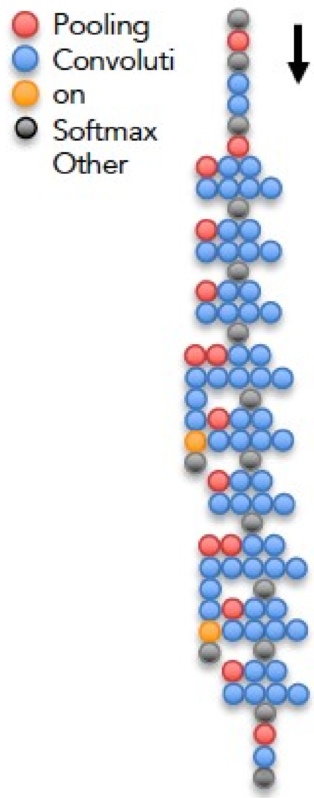
SuperVision



[Krizhevsky NIPS 2012]

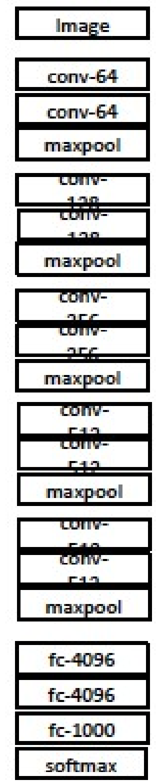
Year 2014

GoogLeNet



[Szegedy arxiv 2014]

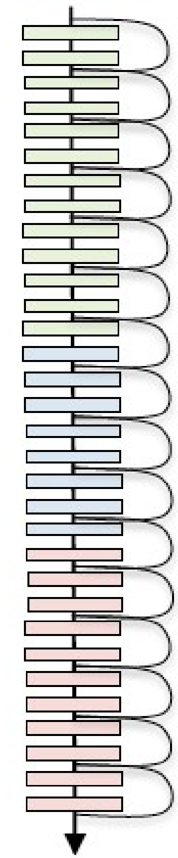
VGG



[Simonyan arxiv 2014]

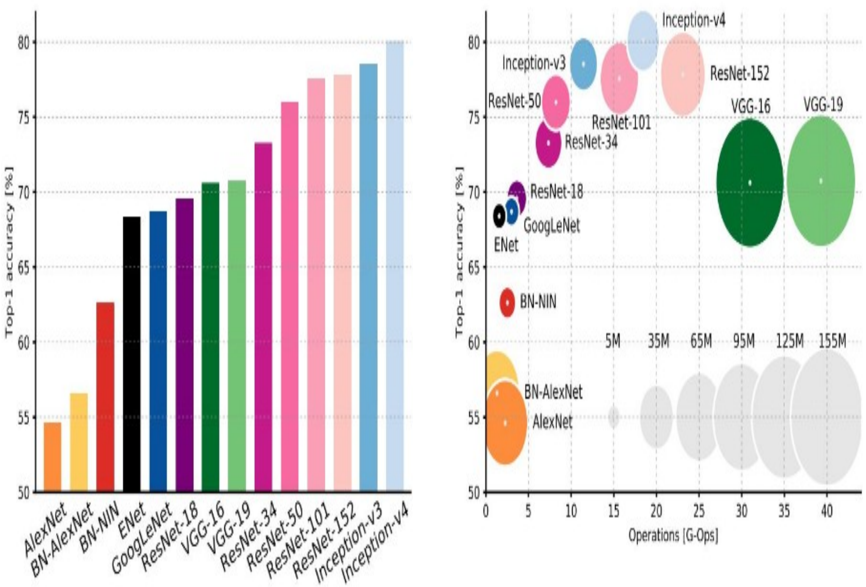
Year 2015

MSRA

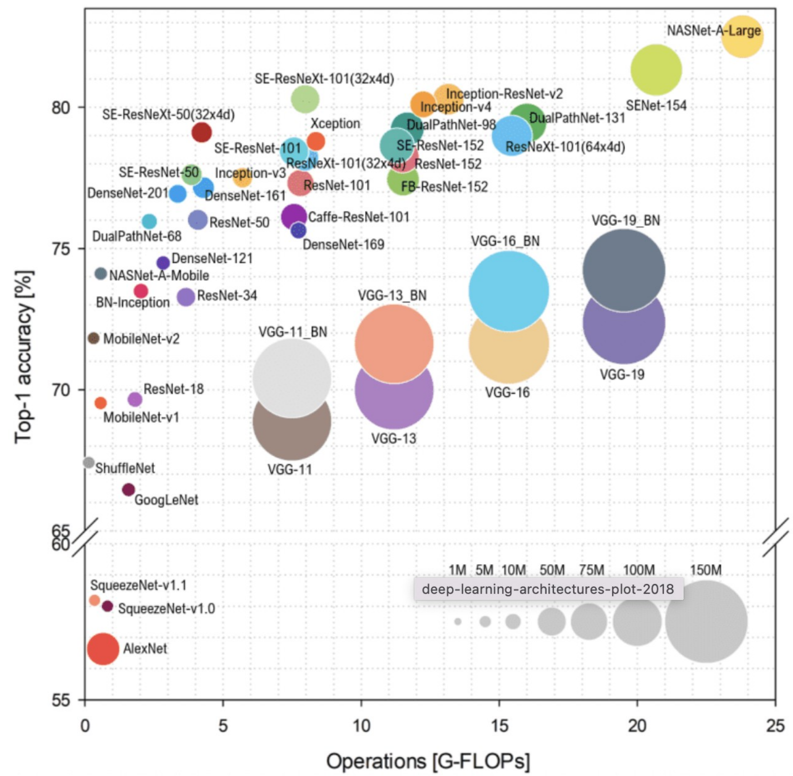


[He ICCV 2015]

Analysis of CNNs



An Analysis of Deep Neural Network Models for Practical Applications, 2017.



- **Millions of parameters!!!**

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

Index

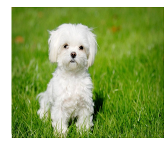
- AI, Machine learning & Deep learning
- What is a Convolutional Neural Network?
  - Layers
  - Loss function and CNN Optimization
- Applications

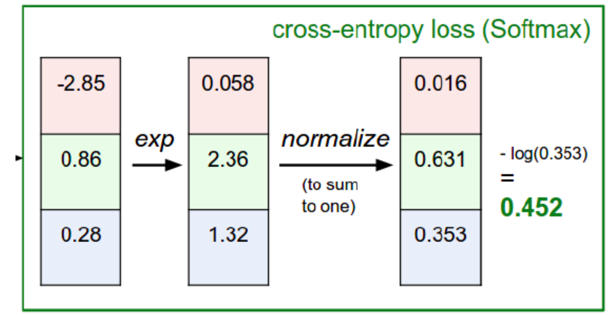
Loss function and optimisation

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

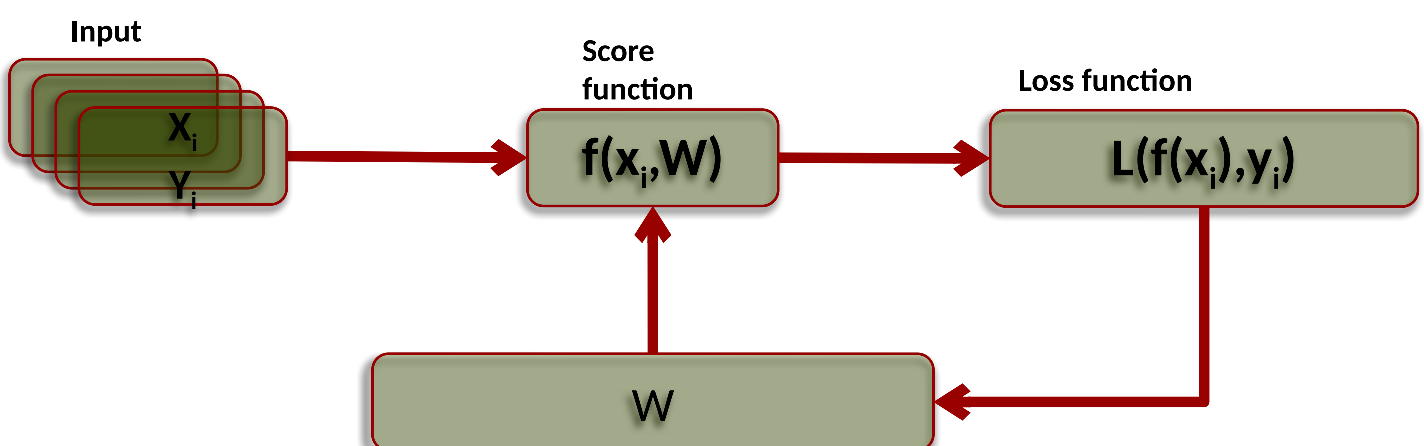
$$L_i = -\log\left(\boxed{\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}}\right)$$

softmax function



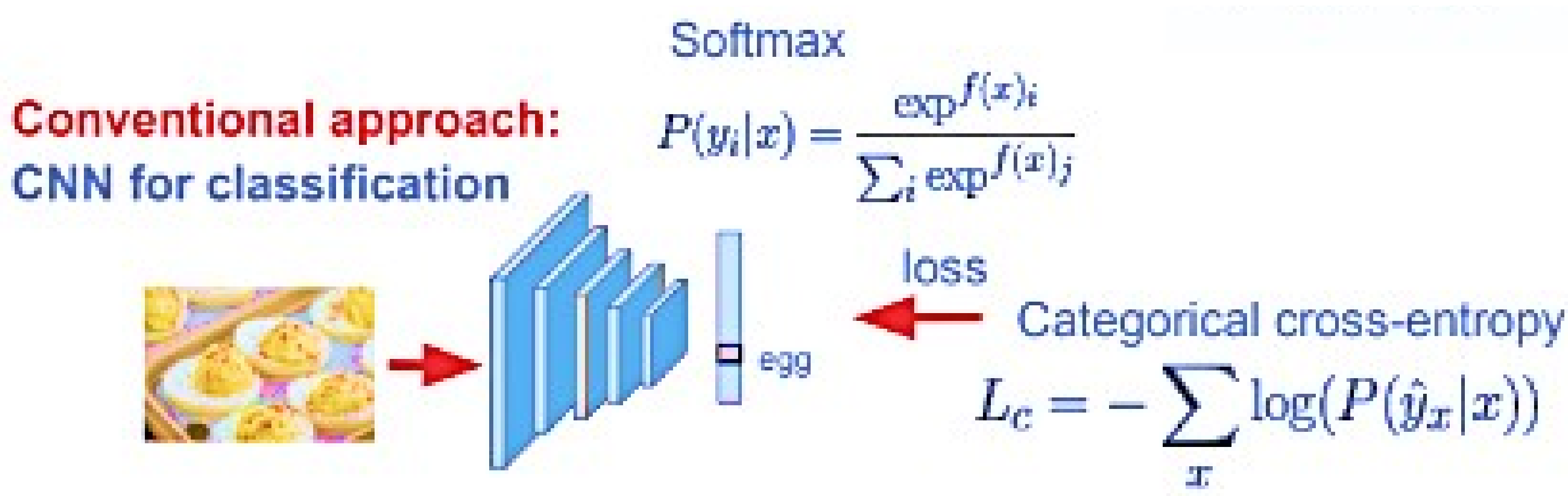


Question : Given the score and the loss function, how to find the parameters W?



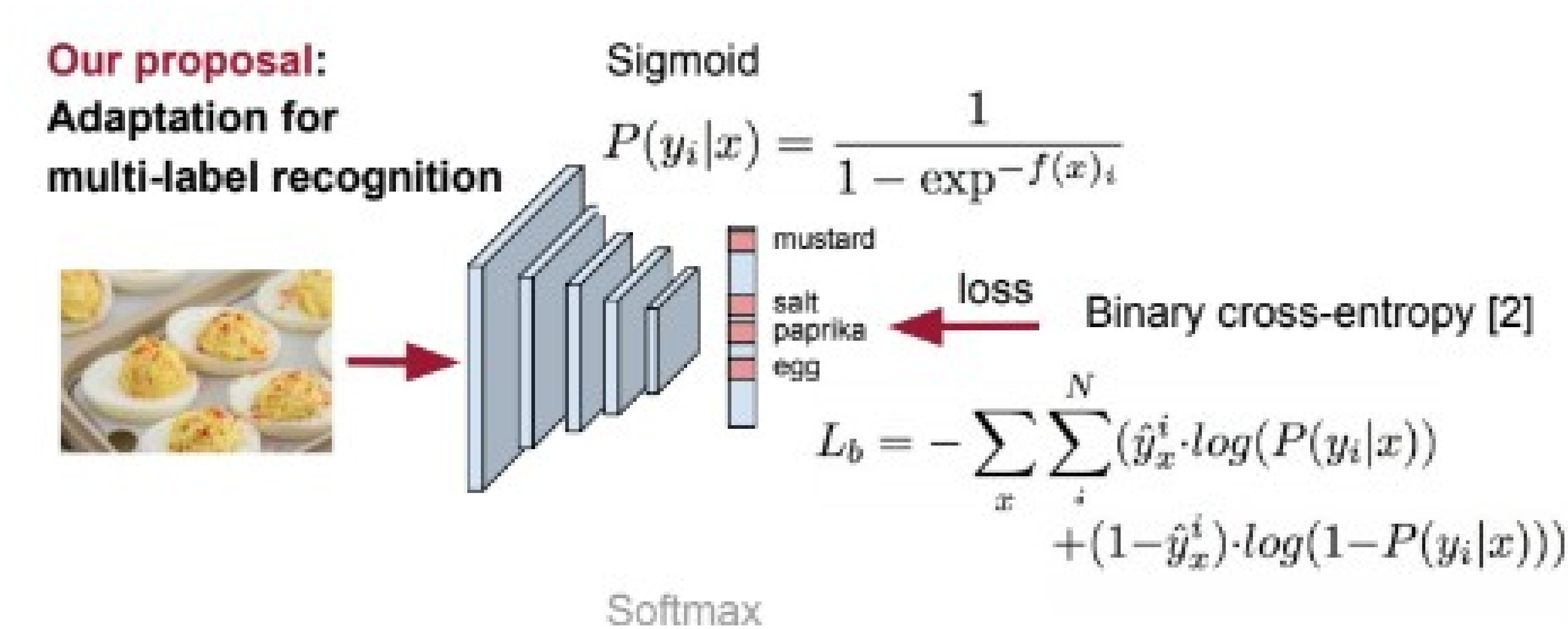
Single-label classification

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

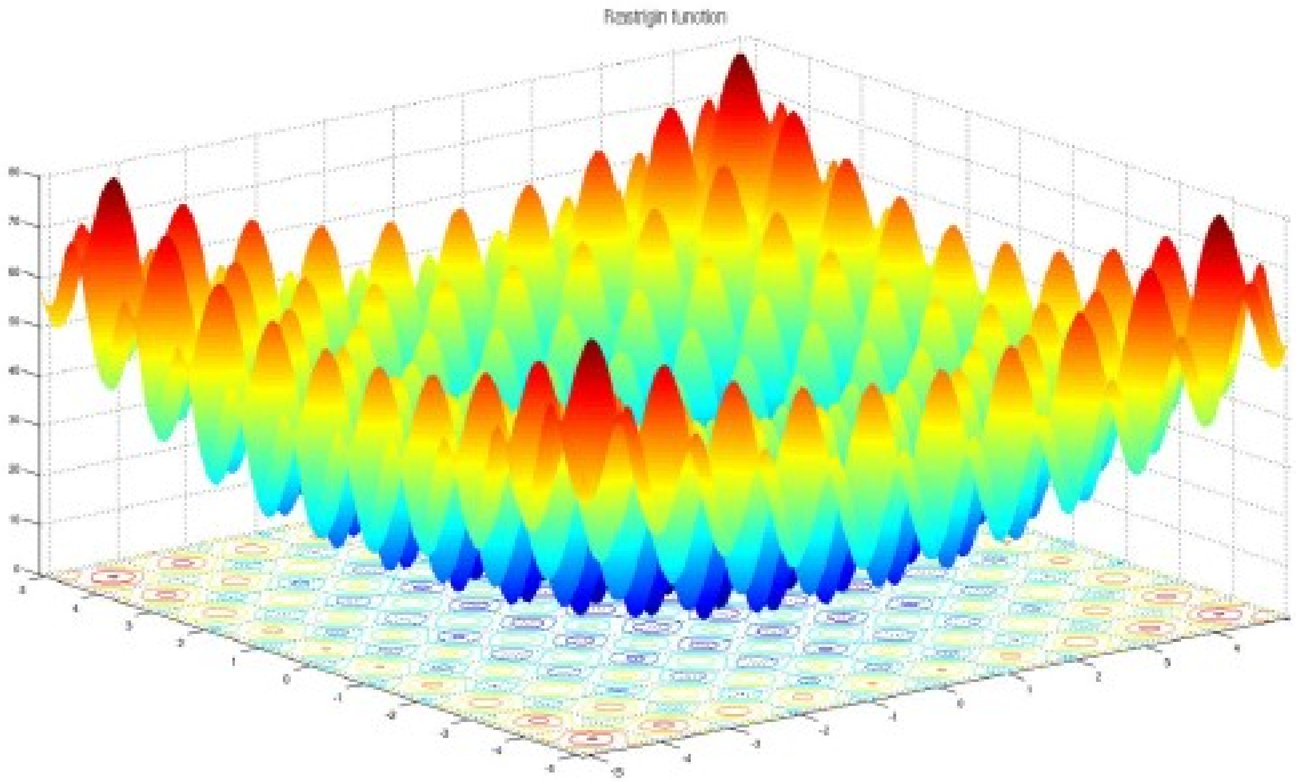


Multi-label classification

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



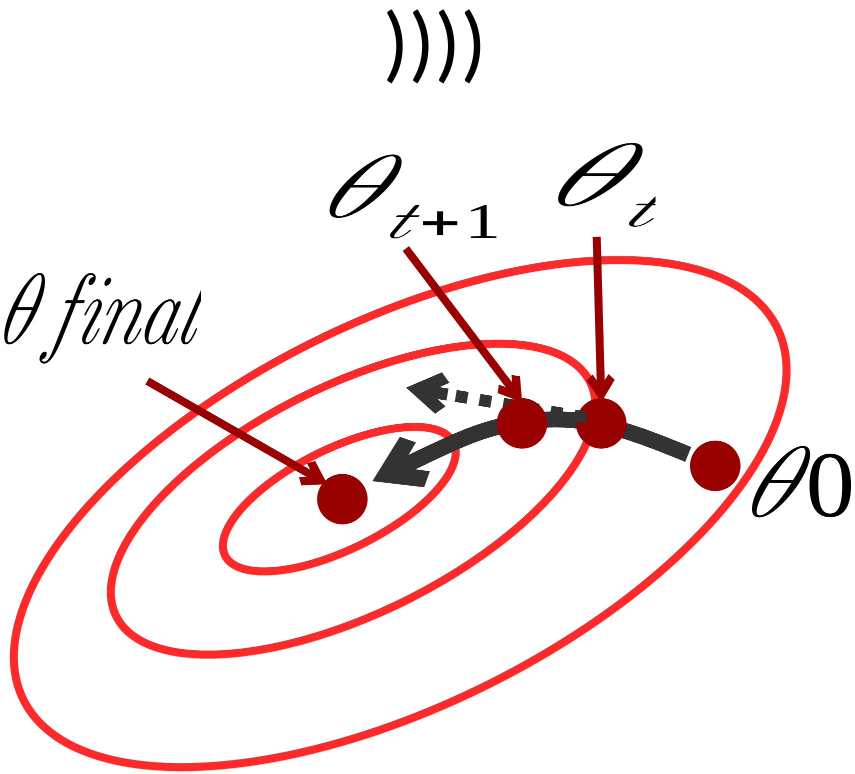
Optimization



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)

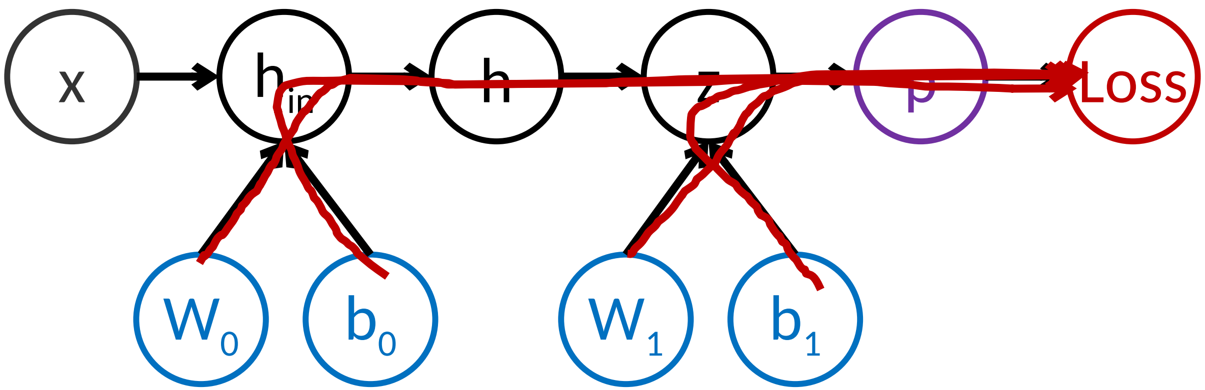
- Computation of requires a full sweep over the training data
- Per-iteration comp. cost = O(n)

Chain rule

- Identify how each variable influence the loss

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

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

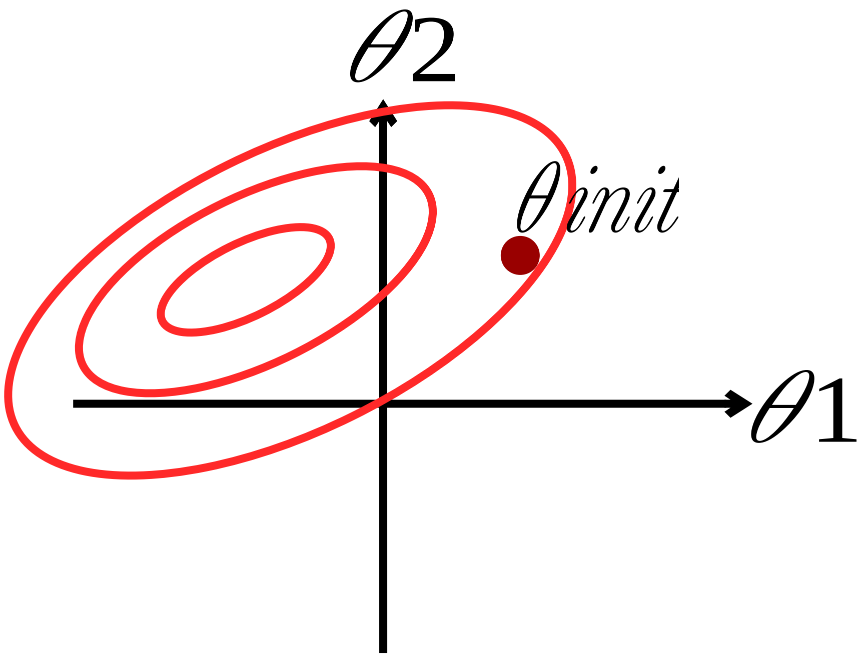


$$\frac{\partial \text{Loss}}{\partial W_0} = \frac{\partial Loss}{\partial p} \cdot \ldots \cdot \frac{\partial \text{h}_{in}}{\partial W_0}$$

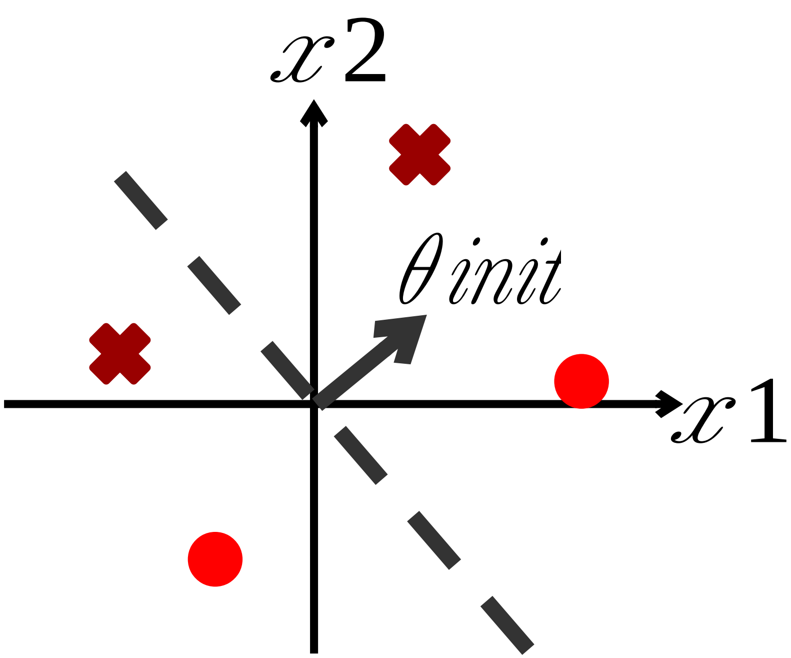
$$\frac{\partial \text{Loss}}{\partial b_0} = \frac{\partial Loss}{\partial p} \cdot \ldots \cdot \frac{\partial \text{h}_{in}}{\partial b_0}$$

Landscape of training objective

Parameter space

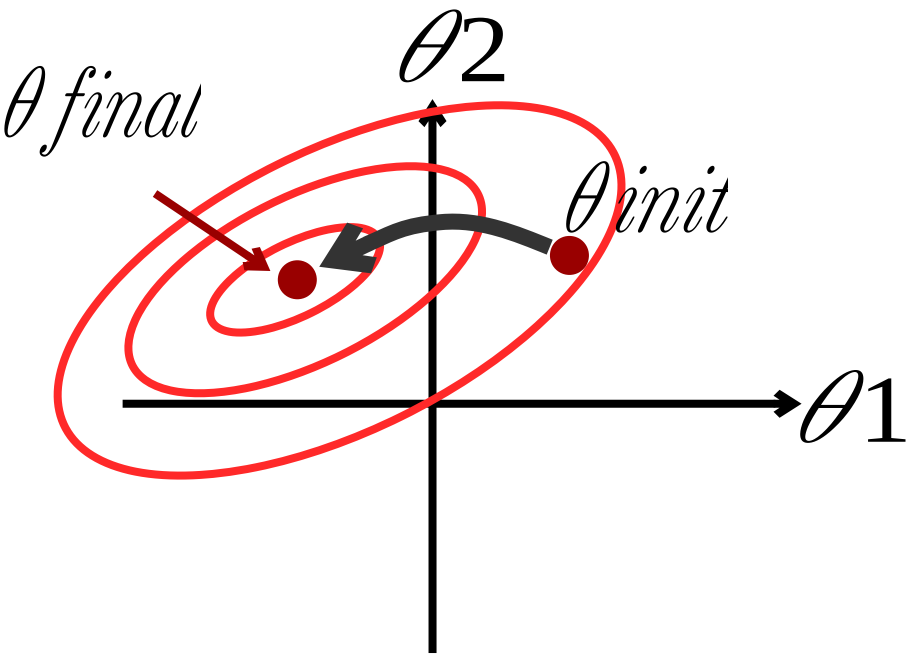


Example space

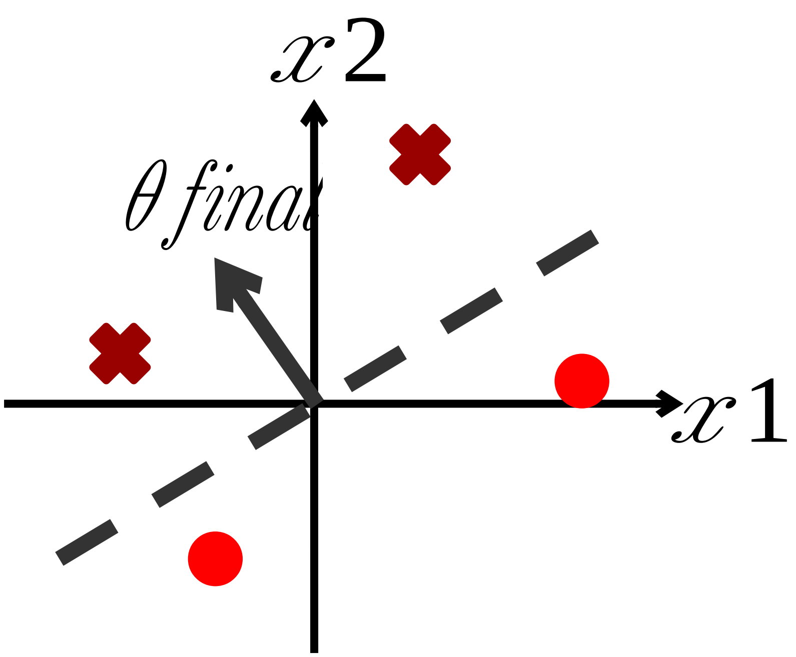


Landscape of training objective

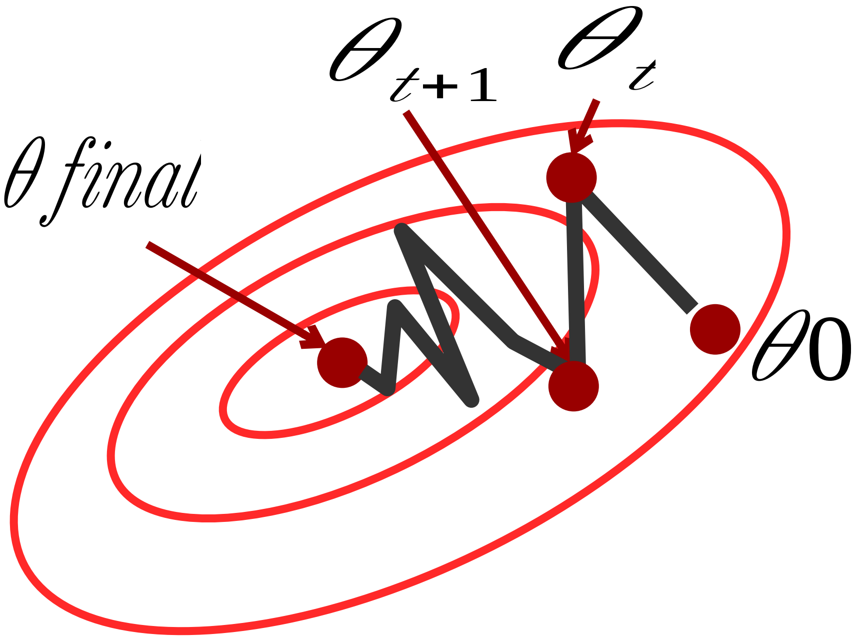
Parameter space



Example space



Stochastic gradient descent (SGD)



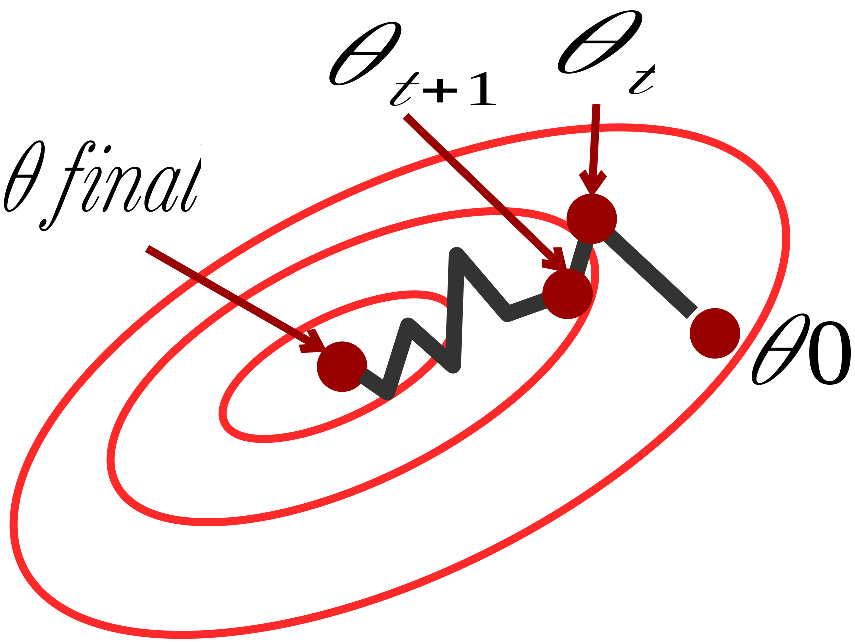
- Initialize randomly
- For t in 0,…, $T_{maxiter}$

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,…, $T_{maxiter}$

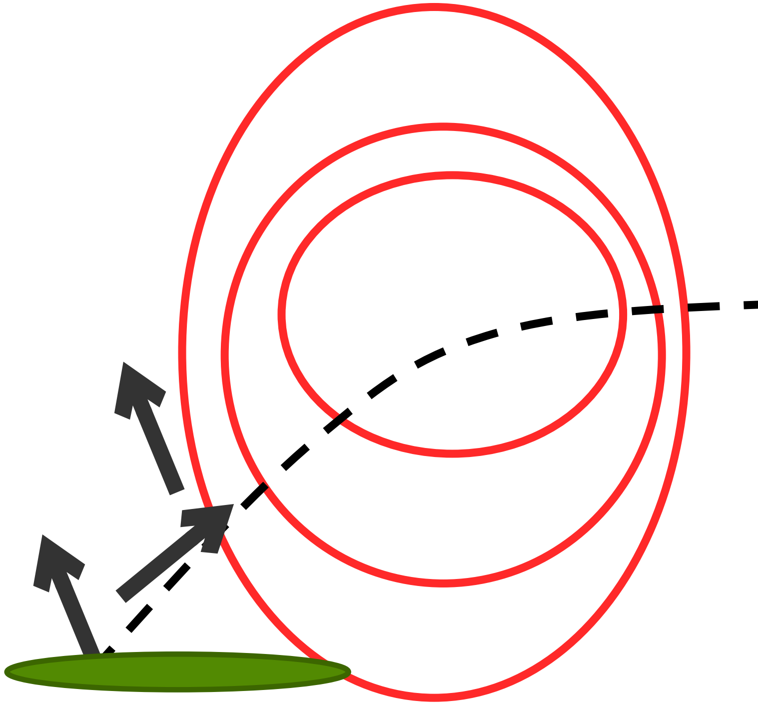
minibatch gradient

where minibatch B is chosen randomly

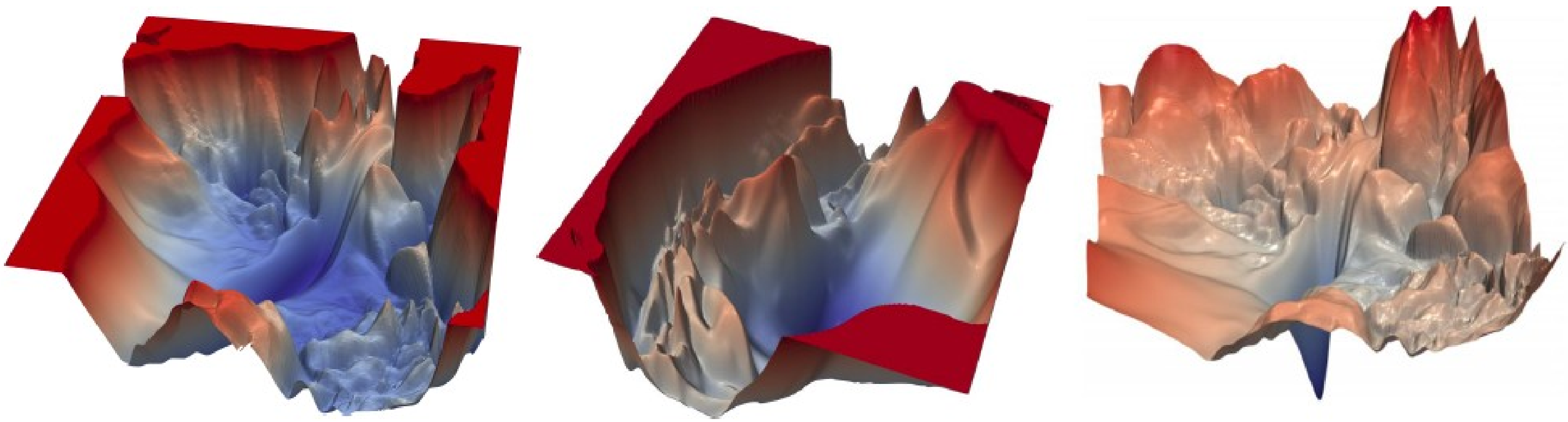
- is average gradient over random subset of data of size B
- Per-iteration comp. cost = O(B)

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



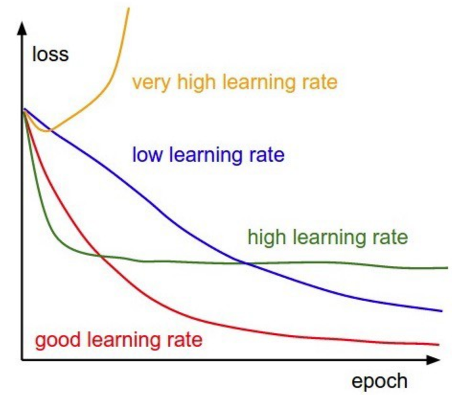
Stochastic Gradient Descent



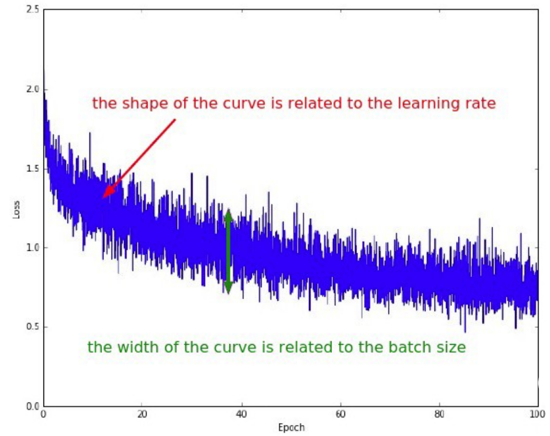
VGG-56 VGG-110 ResNet-56

Hao Li et al., NIPS, 2017

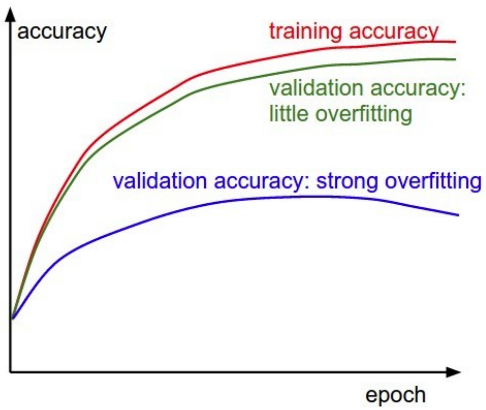
Monitoring loss and accuracy

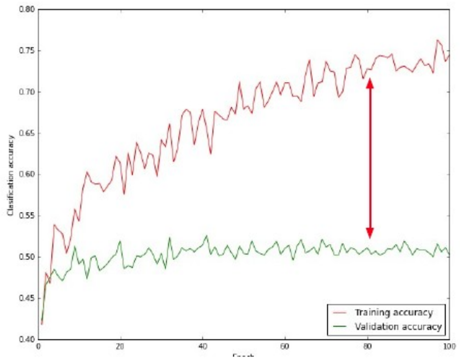


Adapted from Fei Fei slides



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

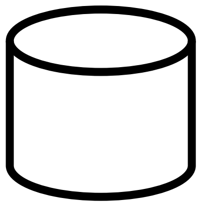




Big gap?
- you're overfitting,
increase
regularization!

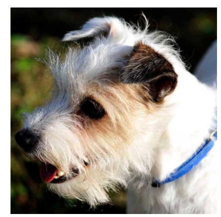
Overfitting – what is signal vs noise?

- Imagine:



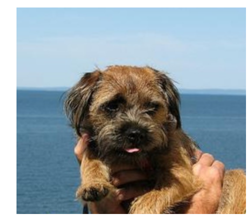
Training data

cat



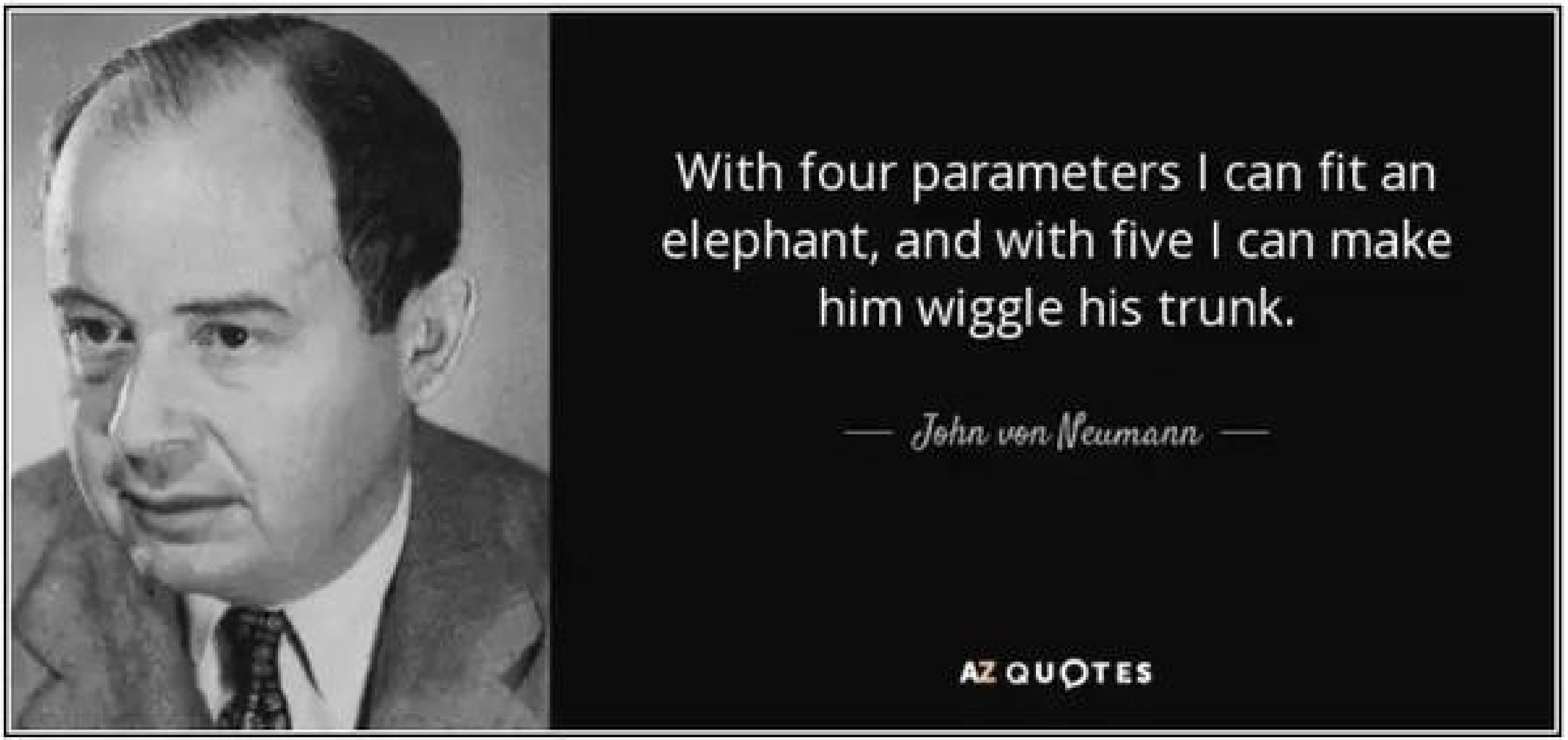
dog

Validation data

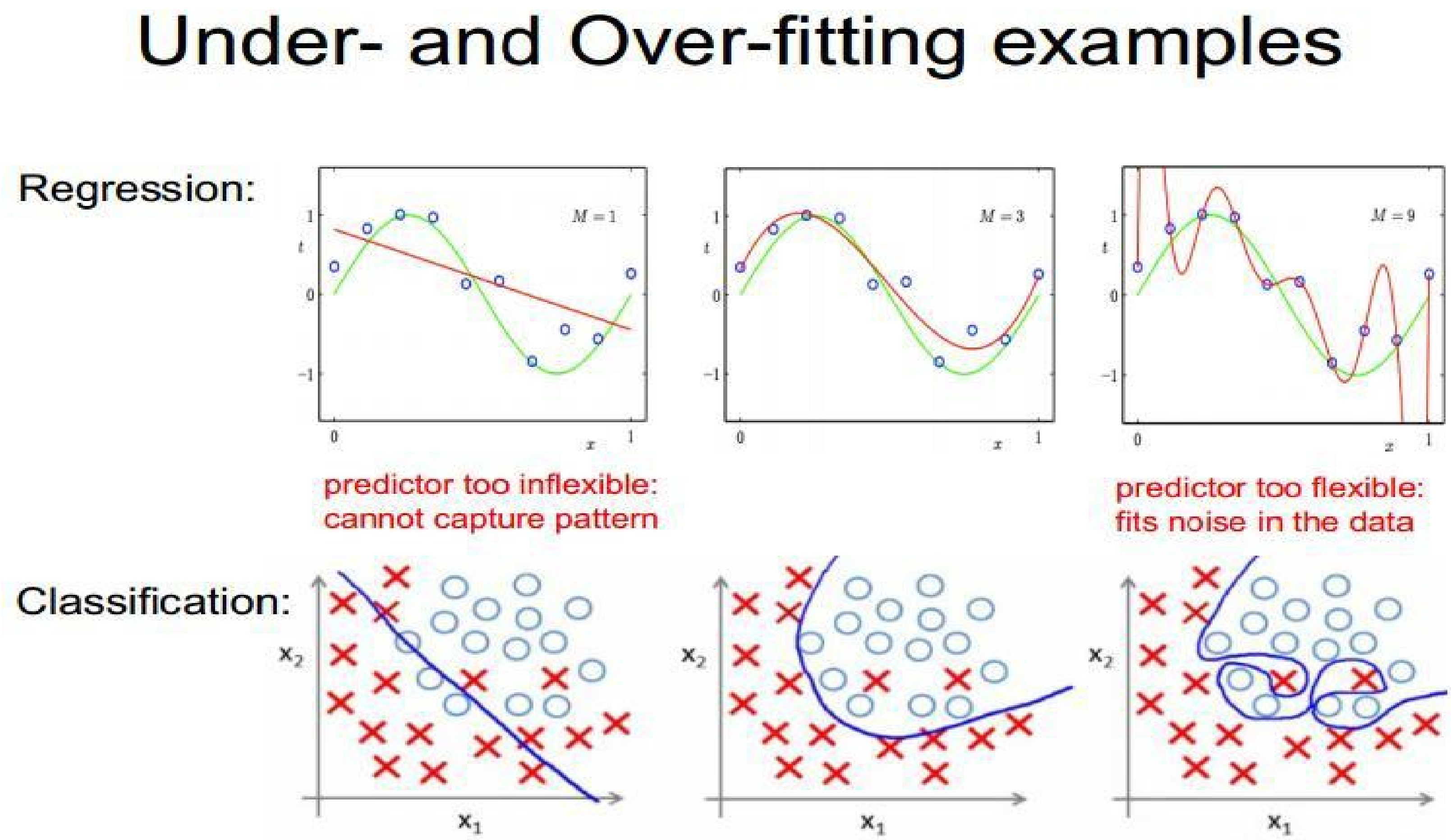


- Powerful models are more likely to overfit
- We need validation data: leave out some portion of the training data to validate the generalizability of the model

Overfitting

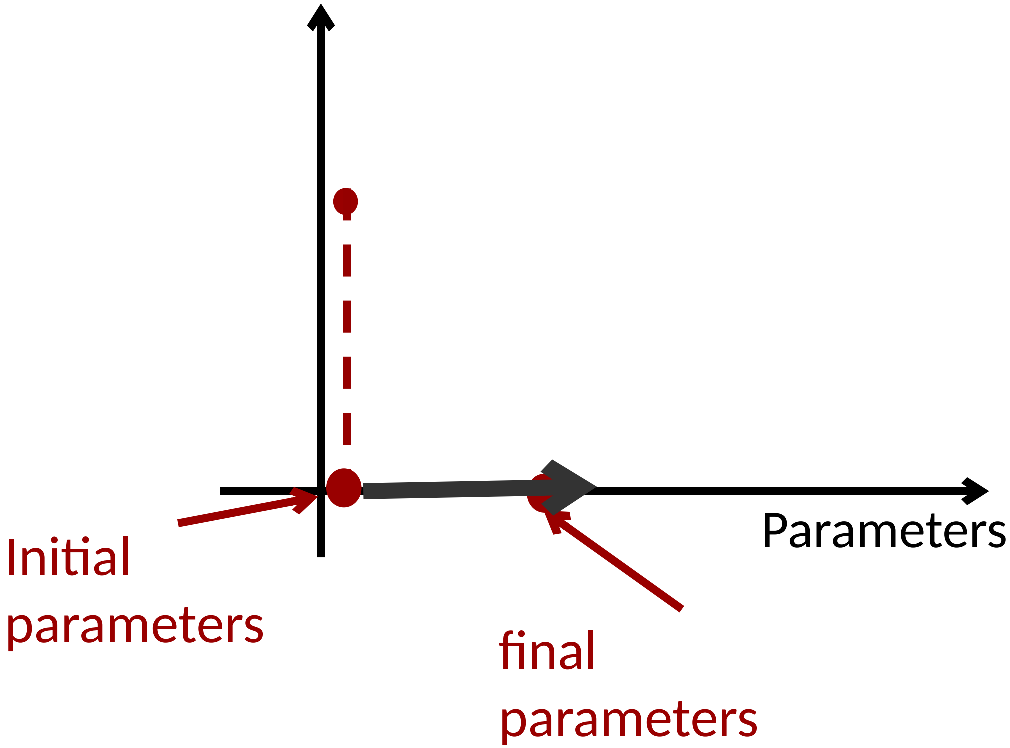


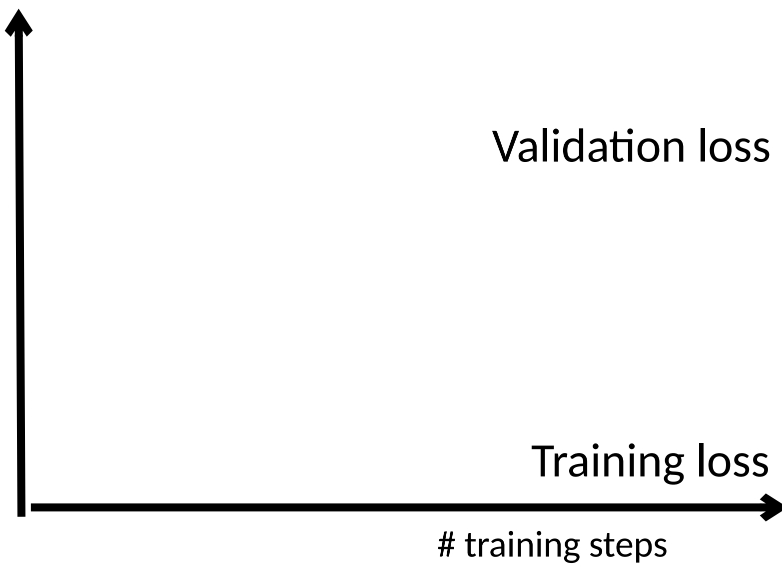
Under- and Over-fitting



Landscape of training objective

Objective function:

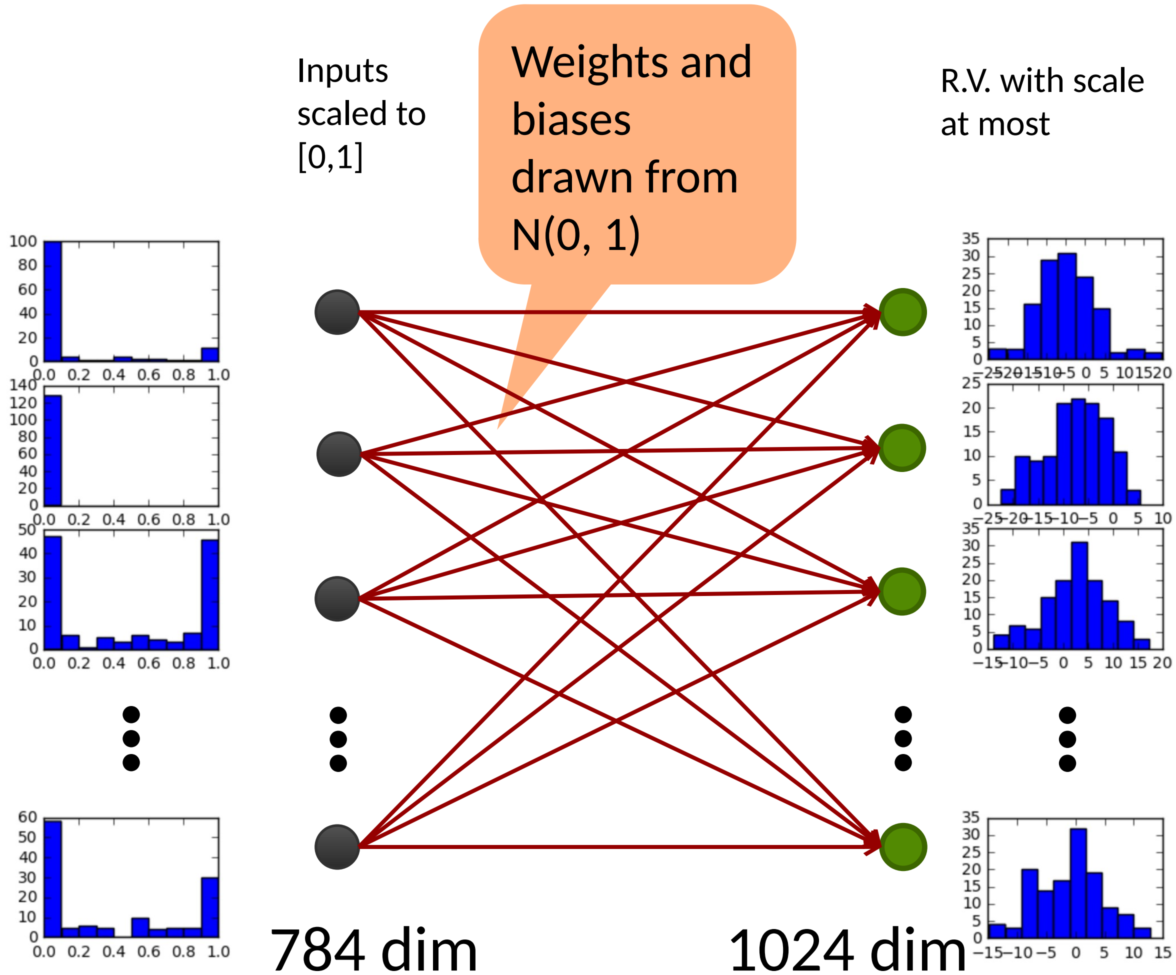




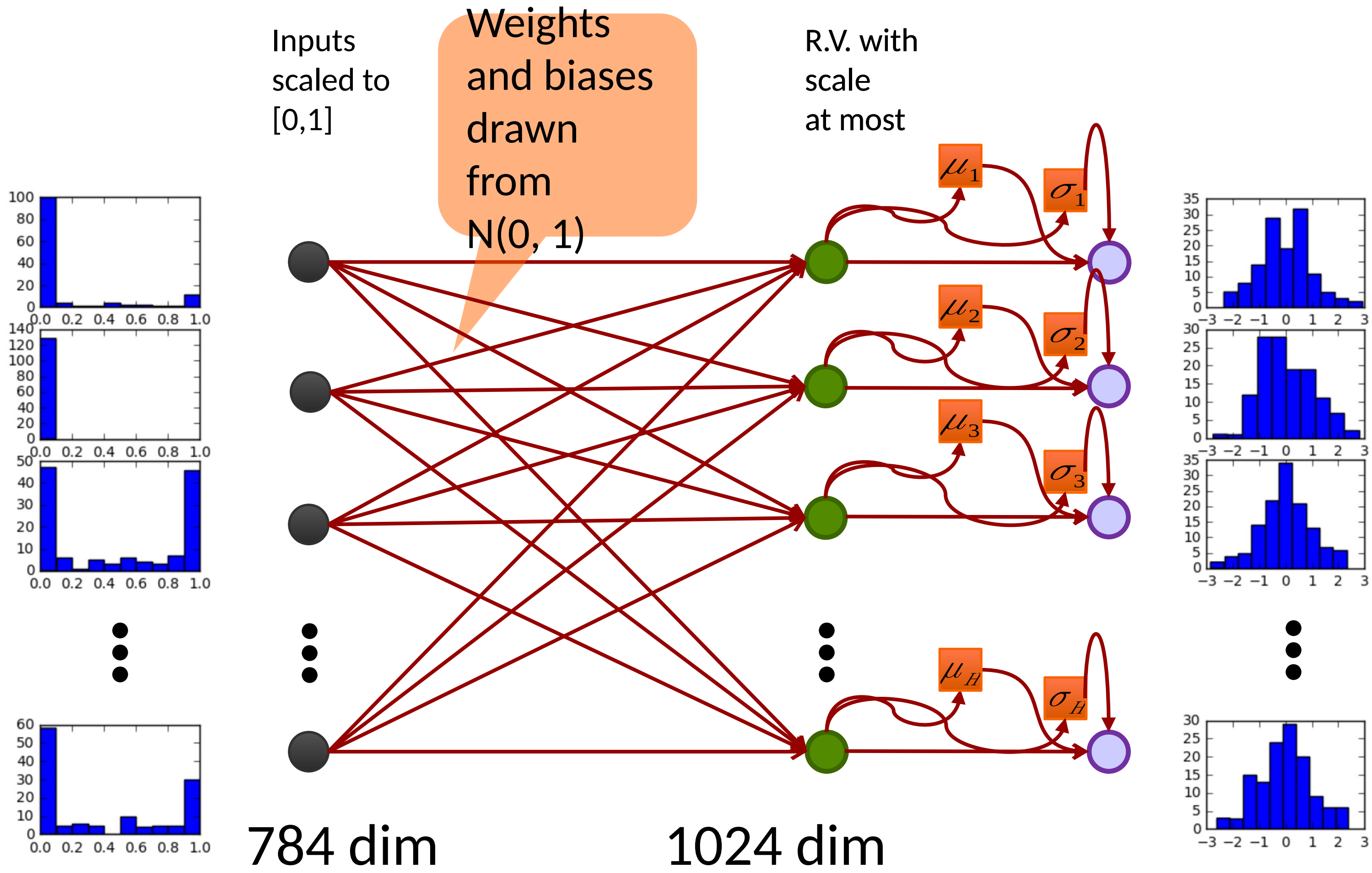
Techniques to reduce overfitting

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

Batch normalization



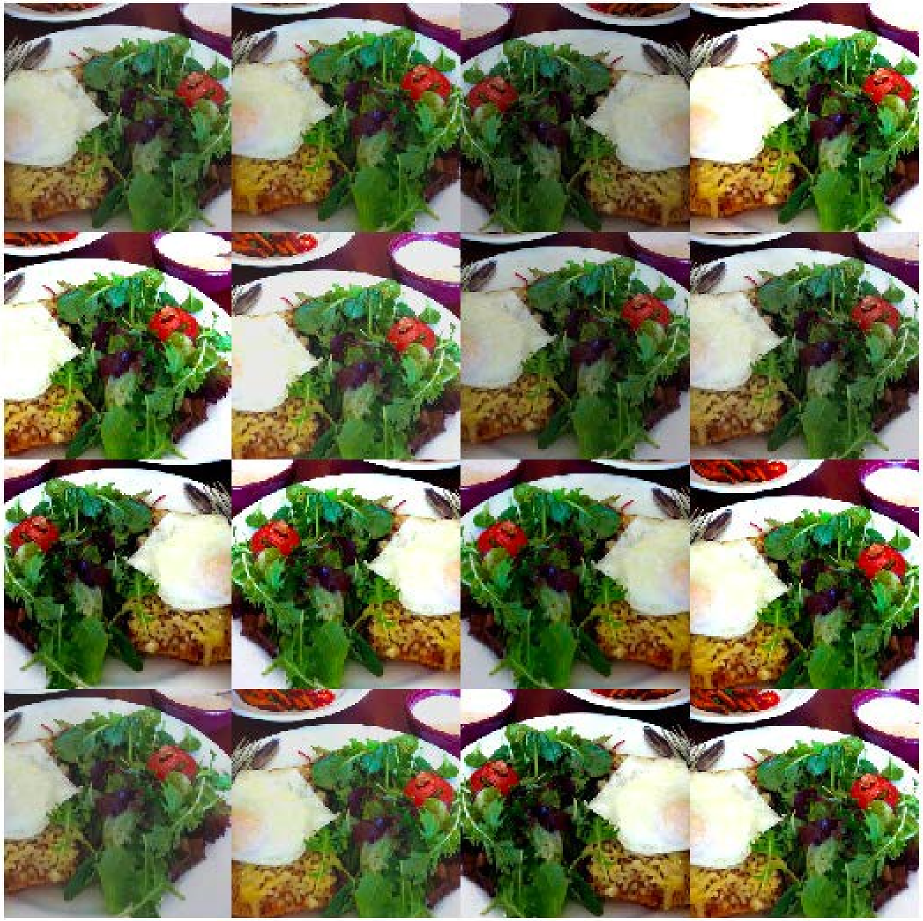
Batch normalization



Benefit: more stable and faster training. Often generalizes better

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

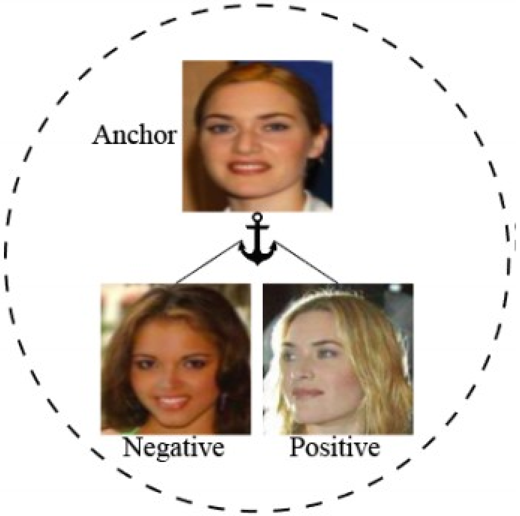
- What is a Convolutional Neural Network?
  - Layers
  - Optimization

- Applications

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.



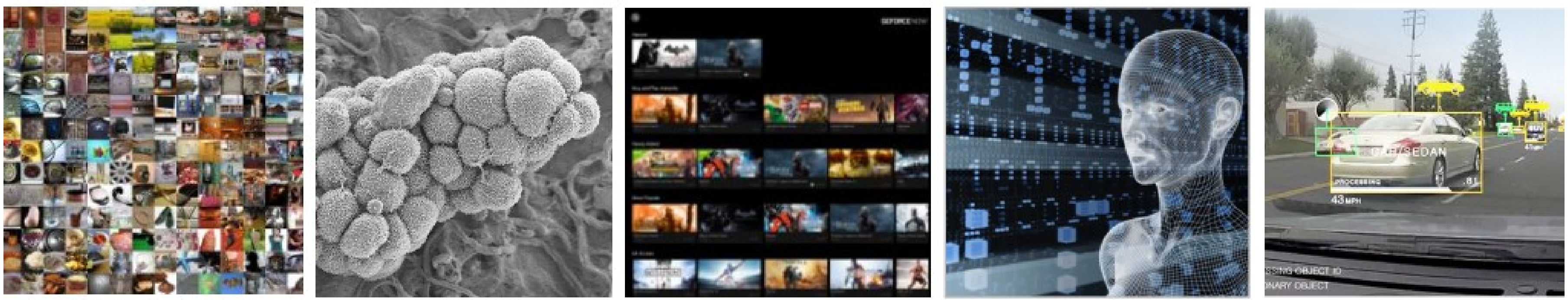
Biometric data for cheating detection I

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

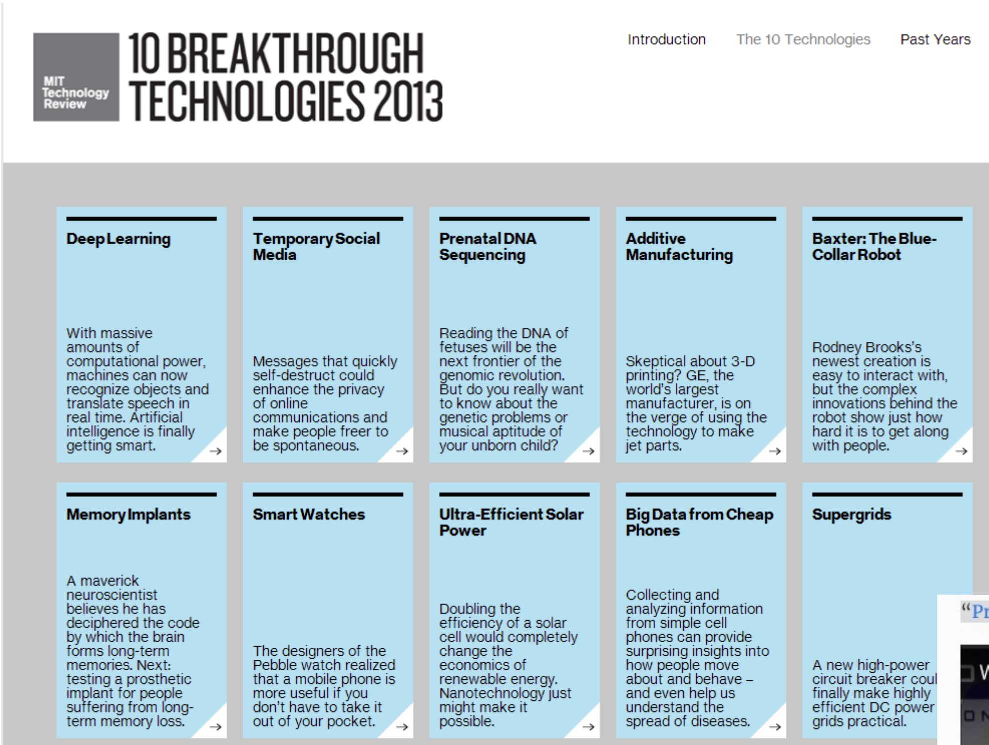
01:42

Deep leearning everywhere



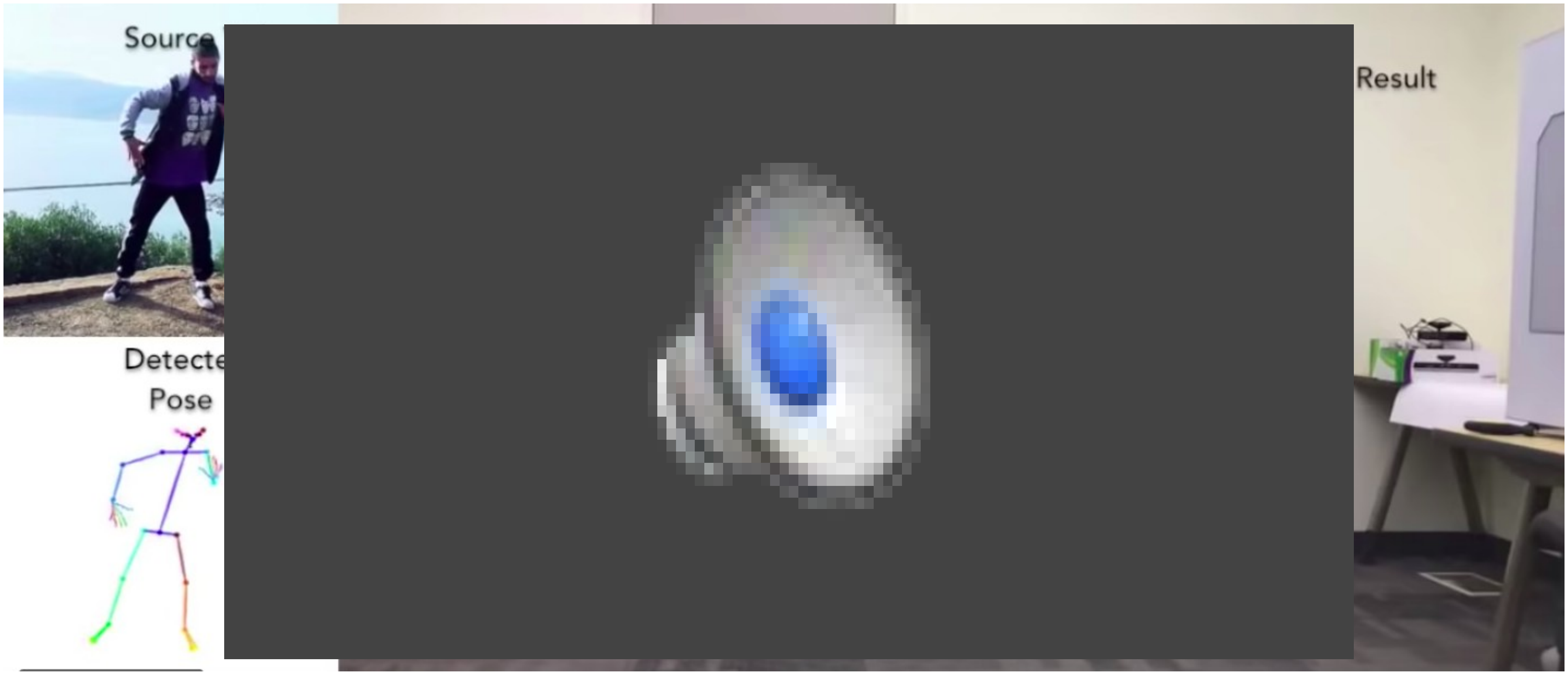
INTERNET & CLOUD	MEDICINE & BIOLOGY	MEDIA & ENTERTAINMENT	SECURITY & DEFENSE	AUTONOMOUS MACHINES
Image Classification Speech Recognition Language Translation Language Processing Sentiment Analysis Recommendation	Cancer Cell Detection Diabetic Grading Drug Discovery	Video Captioning Video Search Real Time Translation	Face Detection Video Surveillance Satellite Imagery	Pedestrian Detection Lane Tracking Recognize Traffic Sign

Deep learning - one of the 10 breakthrough technologies

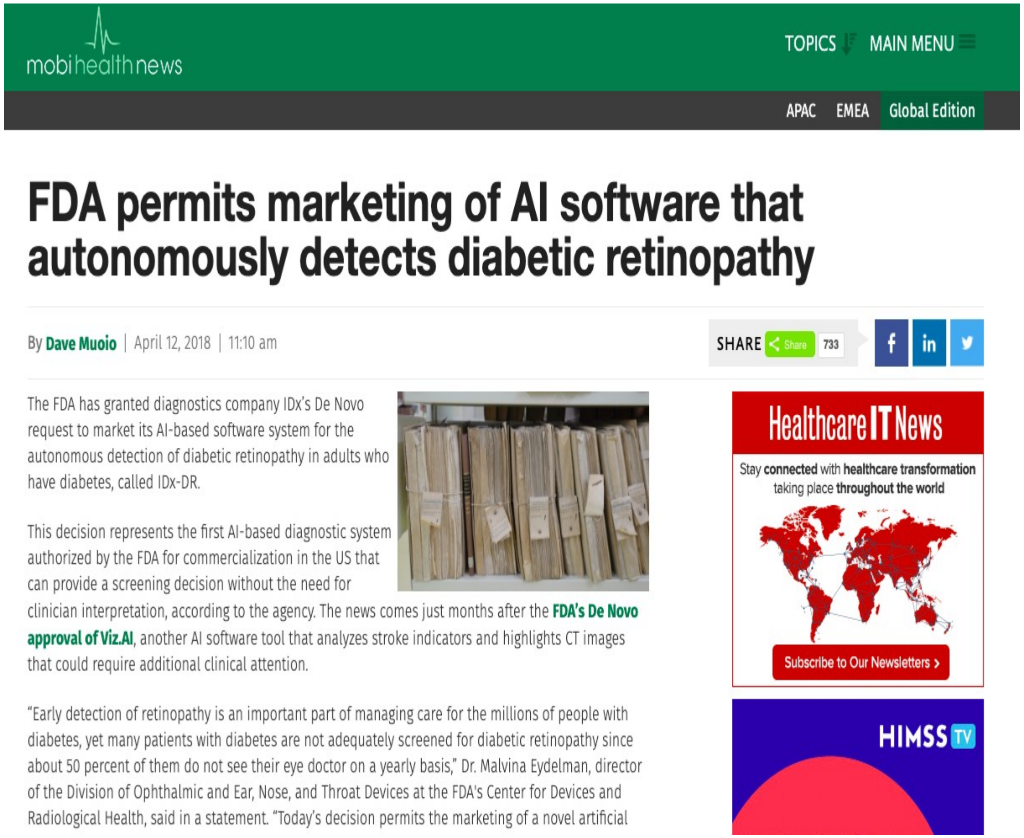


Deep learning everywhere

Everybody dances now



AI & ML: from R+D lab to real life



FDA permits marketing of AI software that autonomously detects diabetic retinopathy

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

SHARE Share 733

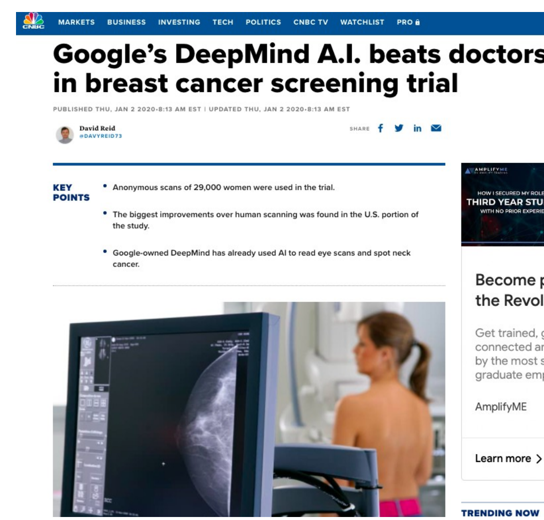
The FDA has granted diagnostics company IDx's De Novo request to market its AI-based software system for the autonomous detection of diabetic retinopathy in adults who have diabetes, called IDx-DR.

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



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

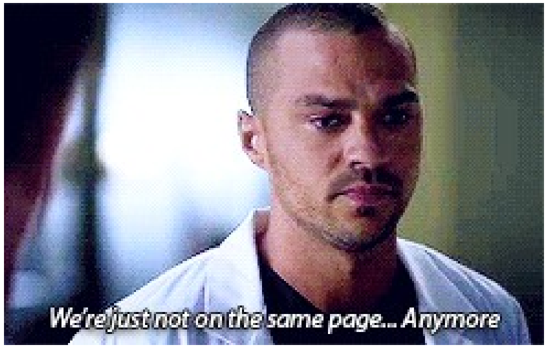
KEY POINTS

- 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.

- AI & ML: from R+D lab to hospital

Concerns about DL

- Different development approach.

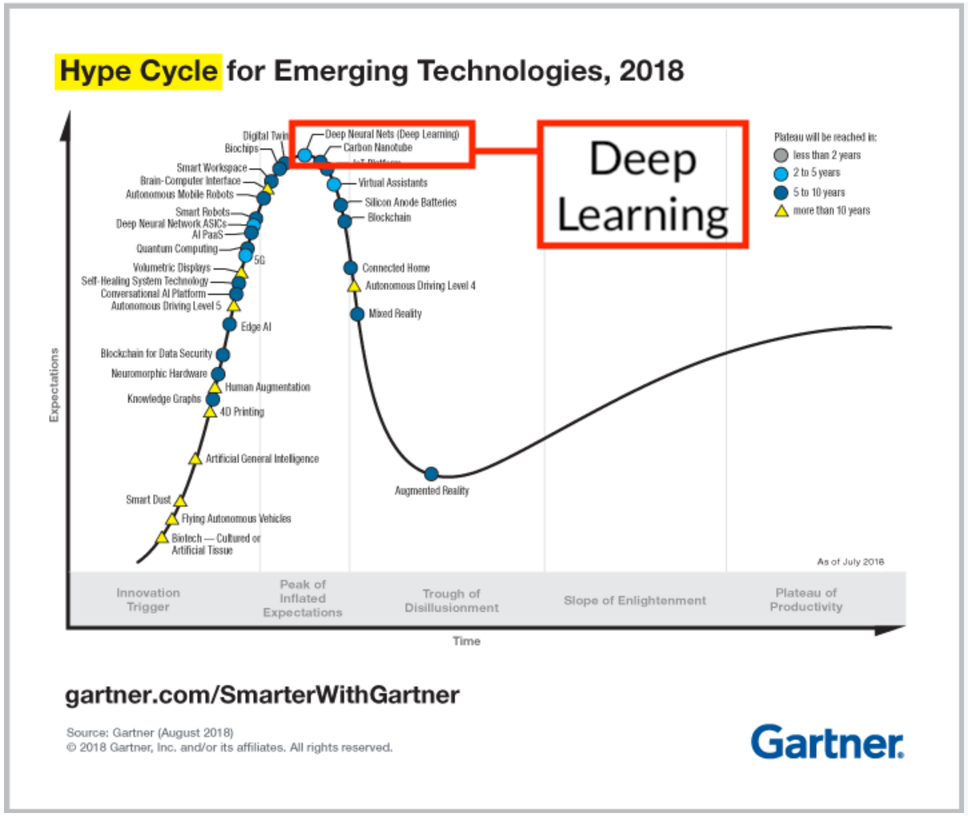


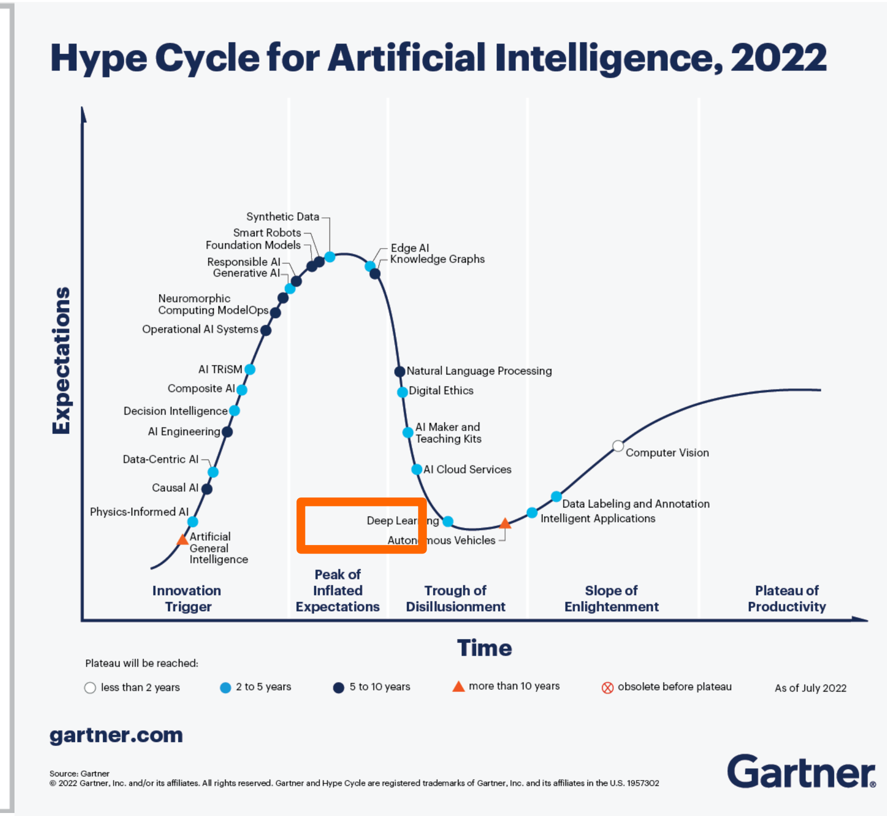
- No clear view on how insight is generated.

- 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

DL and society expectation





[Deep Learning’s ‘Permanent Peak’ On Gartner’s Hype Cycle]

Conclusions I

Deep learning – a technology that came to stay

A (new) technological trend that is affecting directly our environment

Multiple applications

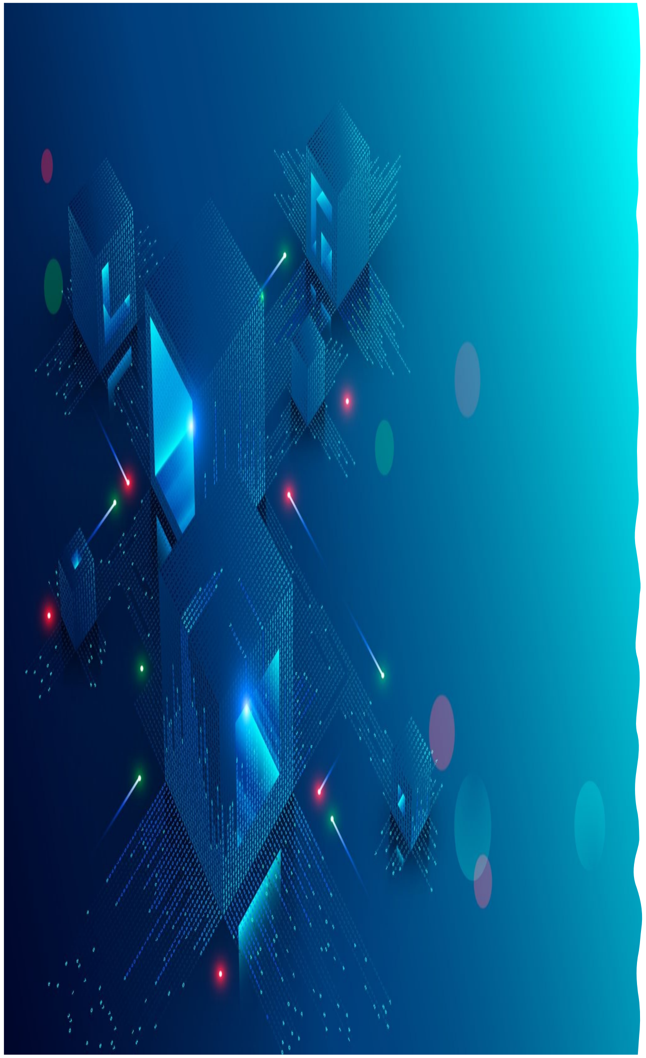
Different CNN models can be found. No optimal one exists.

The most important elements:

- Layers
- Loss function
- Fine-tunning, optimization
- Overfitting

CNN applicable to most Computer Vision problems

Trends in DL



1. **Transfer Learning**: The use of pre-trained models and transfer learning techniques was becoming more widespread, as they can significantly reduce the amount of data required to train effective models.
   - Multi-task learning
2. **Uncertainty modeling:** refers to the process of quantifying and managing uncertainty or ambiguity in the predictions or decisions made by machine learning models
   - Uncertainty-aware MTL
   - Learning with noisy labeling
3. **Self-Supervised Learning**: methods train models on unlabeled data, which can be particularly useful in cases where labeled data is scarce.
4. **Meta-Learning**: Meta-learning techniques were explored to enable models to learn how to learn, which can lead to faster adaptation to new tasks.
   - Continual learning
5. **Generative AI and Generative Adversarial Networks (GANs)**: GANs were being used for a variety of applications, from image generation to data augmentation and domain adaptation.
   - Uncertainty-aware data augmentation
   - NeRFs
   - Stable diffusion

Trends in DL

6. Efficiency and Model Compression: There was a growing interest in making deep learning models smaller, faster, and more energy-efficient, particularly for edge computing and mobile applications.

- Scaling by hierarchical DL models
- Fine-grained recognition

7. Multimodal Learning: Combining information from different sources, such as text and images, to create more comprehensive models for understanding and generating content.

- Oncology, Food ontology

8. Explainable AI (XAI): The need for understanding and interpreting deep learning models became more critical, especially in fields like healthcare and finance where model explainability is crucial.

- Robust Explainable models

9. Federated Learning: This approach allows for training models across decentralized data sources without sharing raw data, preserving privacy. It gained traction, particularly in applications involving sensitive or proprietary data.

10. AI for Healthcare: Deep learning was increasingly applied to medical image analysis, drug discovery, and patient diagnosis, with a focus on improving healthcare outcomes.

- Our projects: food data analysis, omics data,