DATA SCIENCE

CV&IC
COMPUTER VISION

Image classification

Estefanía Talavera Martínez
e.talaveramartinez@utwente.nl
UNIVERSITY OF TWENTE.

Some of the slides are from
Prof Michael Biehl
Prof Petia Radeva
Prof Noah Snavely
Fei-Fei Li, Justin Johnson, Serena Yeung
http://vision.stanford.edu/teaching/cs231n/

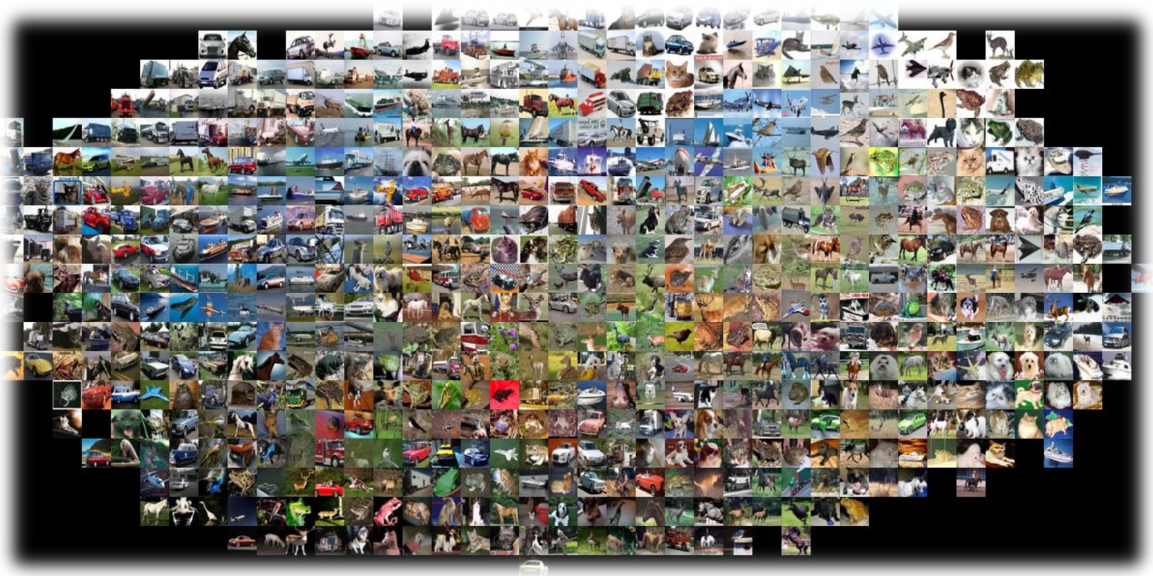
Contents

Brief mention of AI

- Computer vision
- Image classification
    - Machine learning
        - Types of learning
        - Supervised learning
            - (deep)NN and CNNs
            - Regularization and Validation
- Practicals

  Assignments

  Project



“3” CIRI PROJECT

CIRI Project Incident recognition in images

Goal:
Classificaton of incidents depicted in images

Dataset:
Subset of the Incident dataset [1]

Challenges:

- Design and implement ML framework for image classification
- Visualize and represent the data in an informative way
- Evaluate and discuss performance of the implemented models

This session “gave” you tools to work on the project from different perspectives, Investigate them further and use them to build something nice!

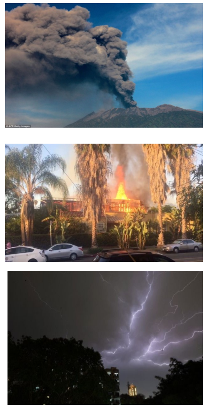
[1] Weber, Ethan, et al. "Detecting natural disasters, damage, and incidents in the wild." *European Conference on Computer Vision*. Springer, Cham, 2020.

CIRI Project Image classification framework

Input data

Images (+labels) + supervised ML = Image classification

…



Feature extraction → Feature processing → ML

Classification

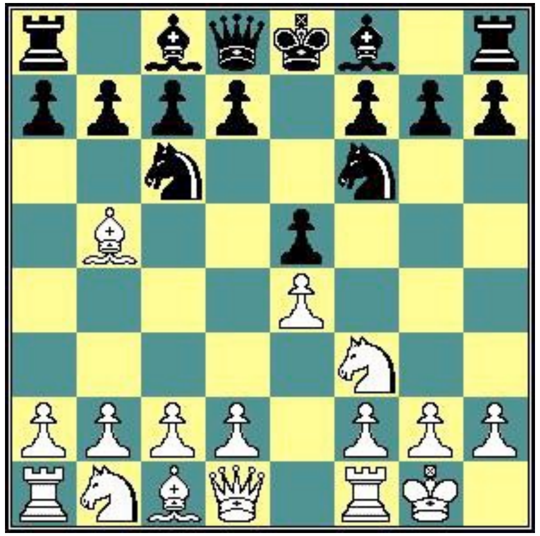
- **0** Flooded
- **1** Airplane accident
- **2** Oil spill
- **3** Burned
- …
- **n-1** Thunderstorm
- **n** Volcanic Eruption

Images: Weber, Ethan, et al. "Detecting natural disasters, damage, and incidents in the wild." *European Conference on Computer Vision*. Springer, Cham, 2020.

0 BRIEF MENTION OF AI

Can *machines* (automata, computers, programs) be intelligent ?

… think ?

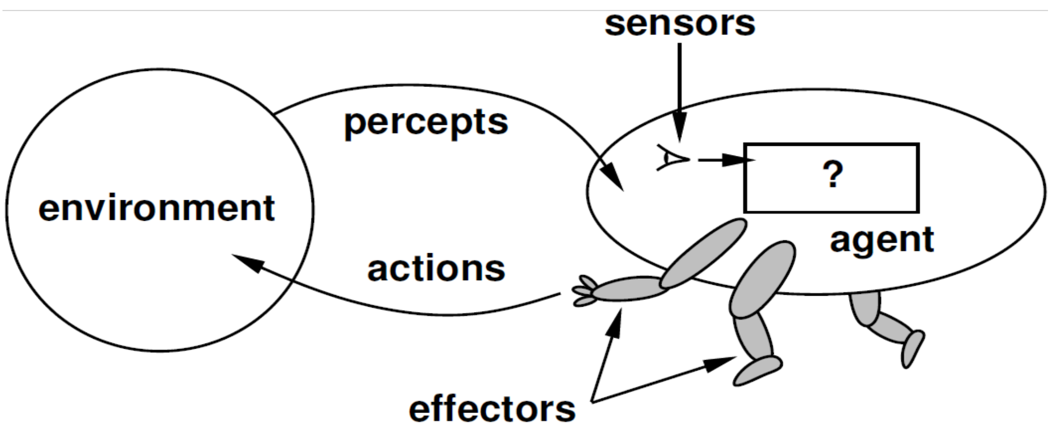


Computer chess

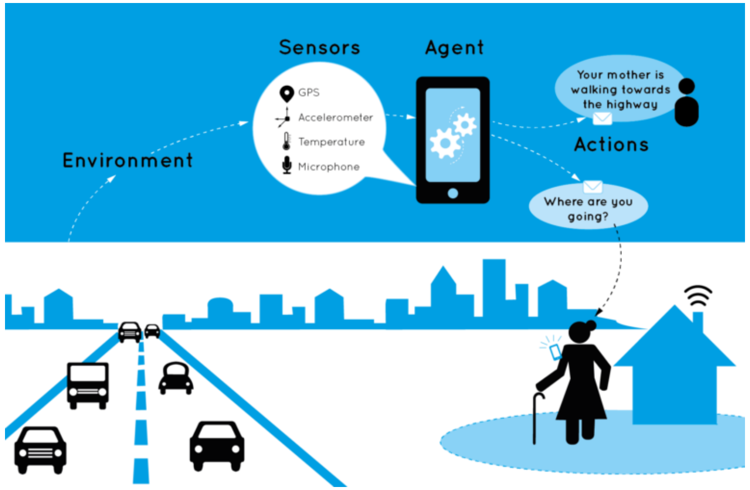
- Deep Blue beats Kasparov (May 1997)
- Matches expert level performance
- 'Thinks' differently from human expert …
  by examining ~ 200 million possible situations

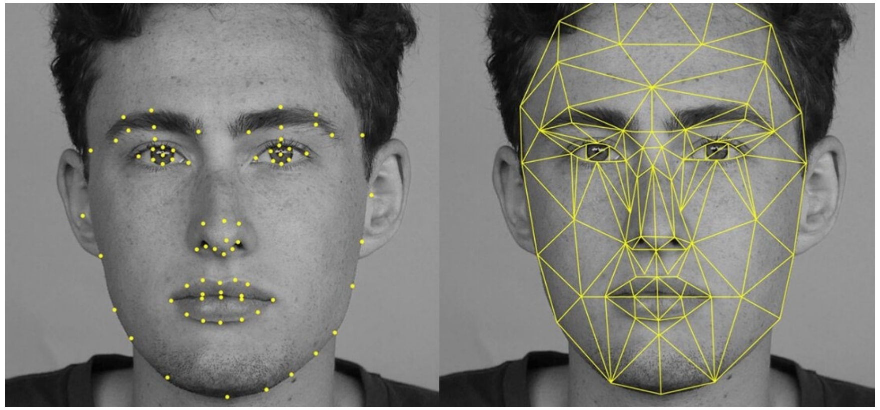
real intelligence or *"just computation"* ???

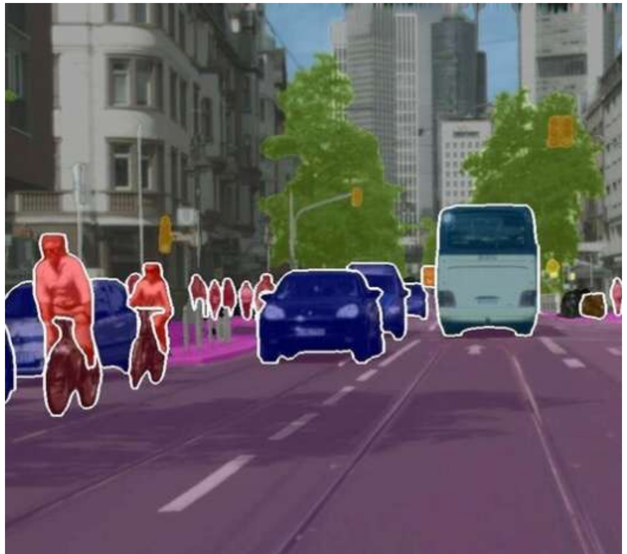
Software-based AI systems

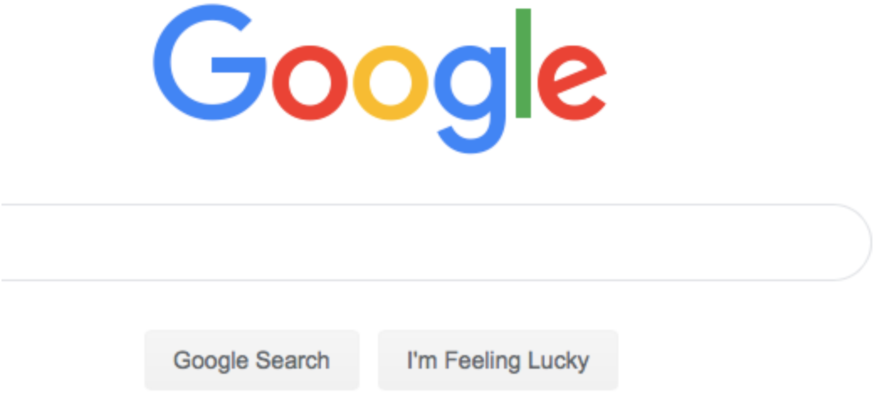


Software-based AI systems









Slide Birna vanRiemsdijk

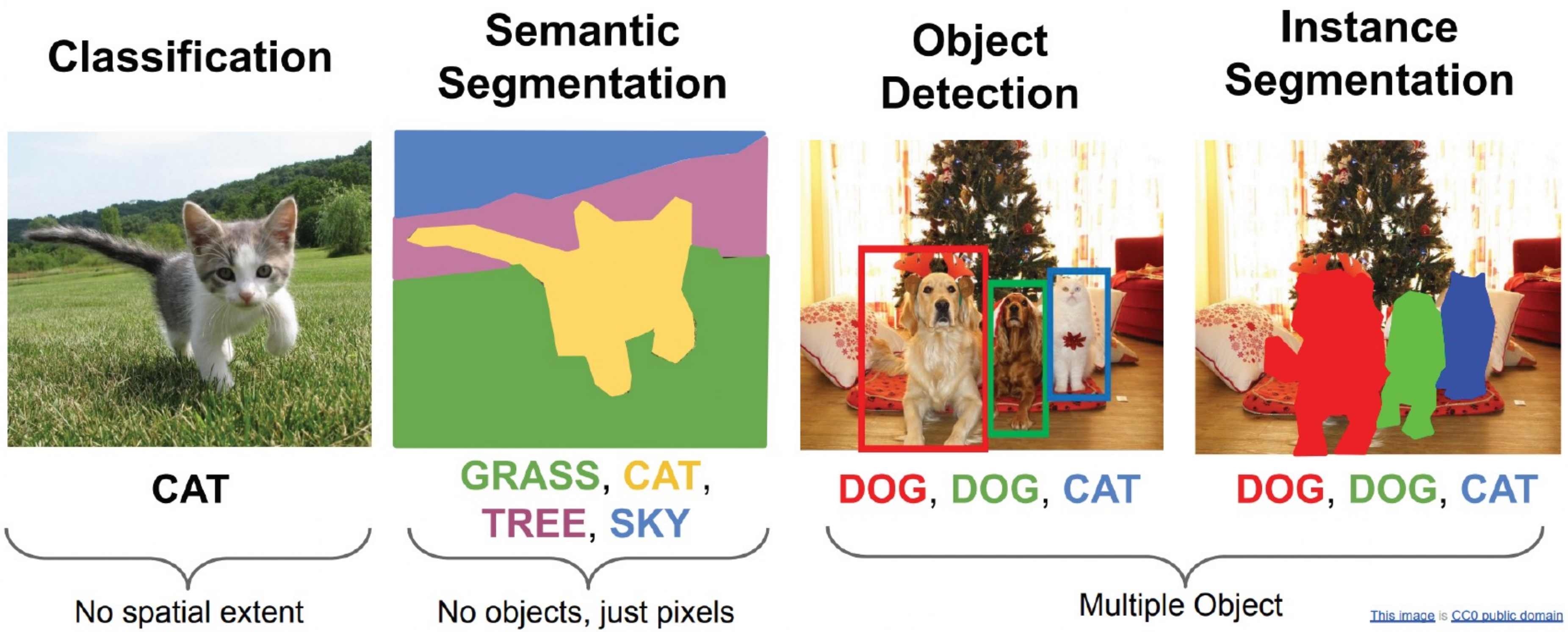
1 COMPUTER VISION

Computer vision

Make computers understand images and video.

- What kind of scene?
- Where is the cars?
- How far is the building?
- How many people are present?
- ...

Computer vision *tasks*



slide from http://cs231n.stanford.edu

Computer vision *tasks*

Input: **x**

Is this an urban or rural area?

Output: **y = { -1, +1 }**

Binary Classification

Which city is this?

Output: **y = { 0, 1, 2, …, C }**

Multi-class Classification

Computer vision *tasks*

Input: **x**

Where is the object in the image?

Output: **y = { x, y, w, h } coordinates**

Object Detection

What is the semantic class of each pixel?

Output: **y ={ 0, 1, 2, …, C } per pixel**

Semantic Segmentation

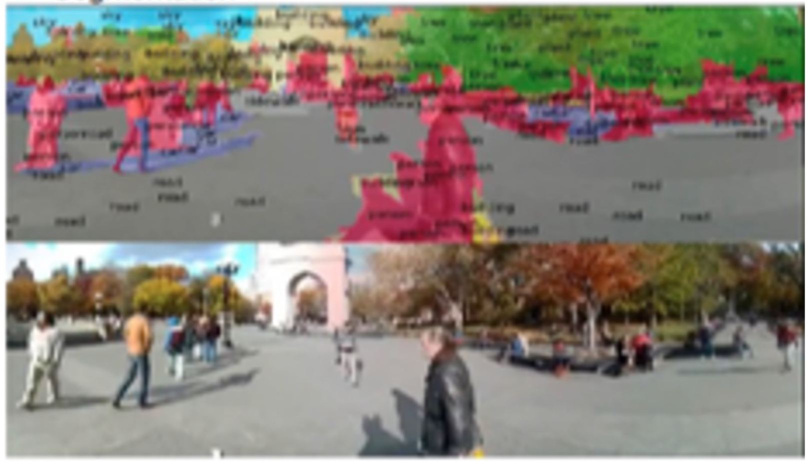
Computer vision *tasks*

Classification

Retrieval

Detection

Segmentation



[Faster R-CNN: Ren, He, Girshick, Sun 2015]

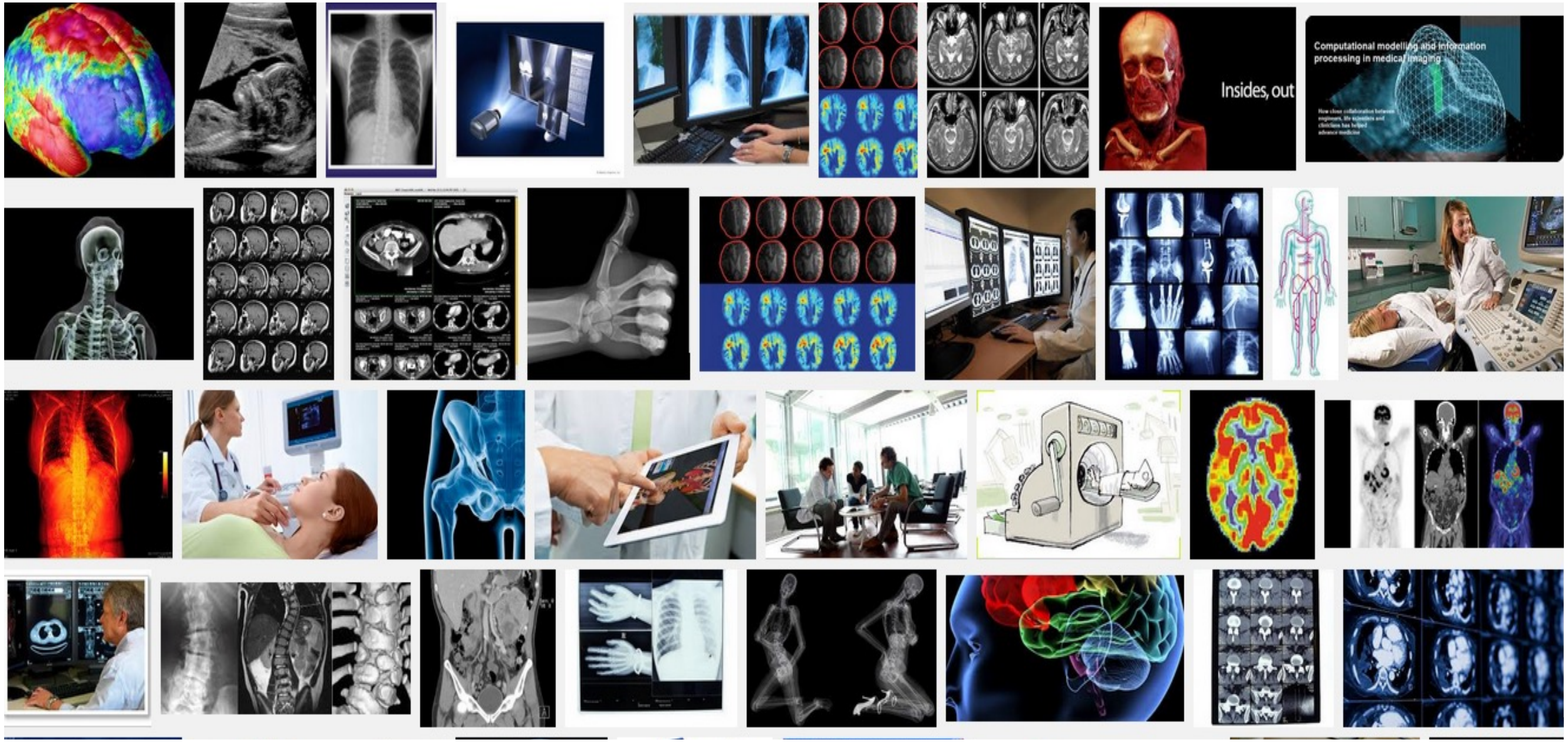
[Farabet et al., 2012]

Computer vision *tasks*

reddit.com/r/deepdream

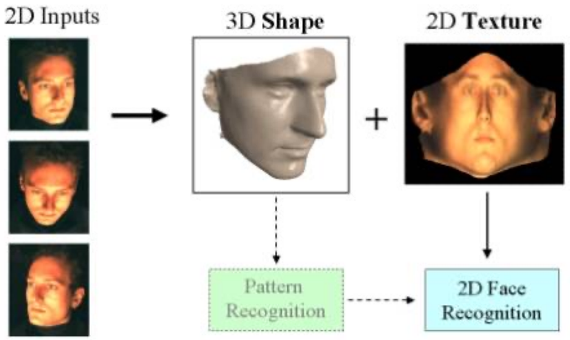
[Vinyals et al., 2015]

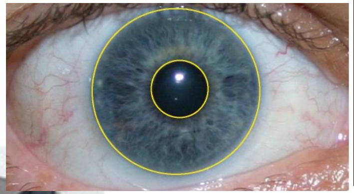
Applications
Medical images and health



Applications
Surveillance and security: biometric systems

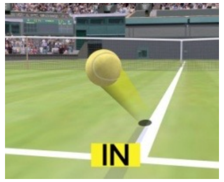
- Fingerprint identification, facial recognition, iris and retinal scan, hand geometry, geometry of the ear, signature recognition, voice identification, identification of the DNS, the smell of human characteristics or recognition of typing motion (not a privilege of science fiction any more).



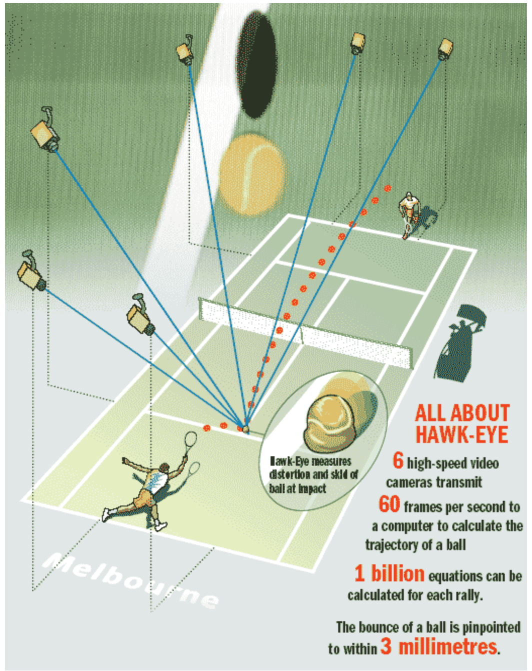


Applications
Media and entertainment

- Augmented TV



- Hawk Eye: tracking the ball in game sports.

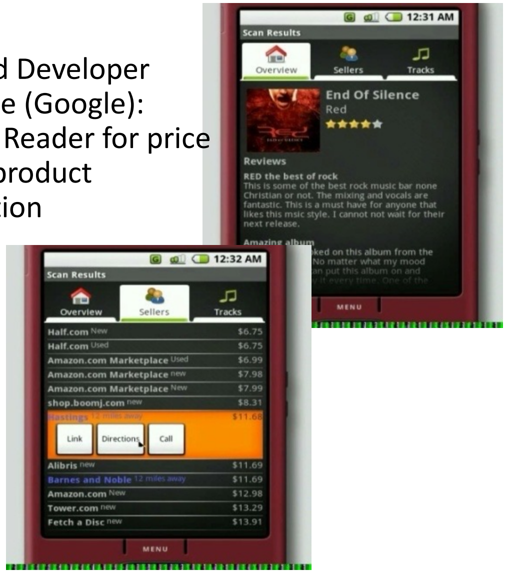


- PVI: Virtual publicity in real television pictures.

Applications
Mobile computer vision

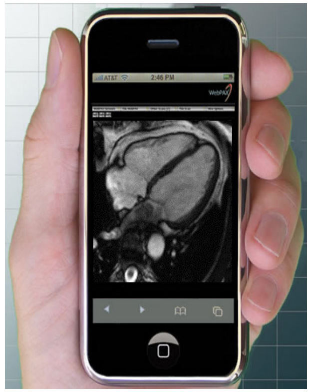
•Client services

•Android Developer Challenge (Google): Barcode Reader for price list and product information



•For visually impaired

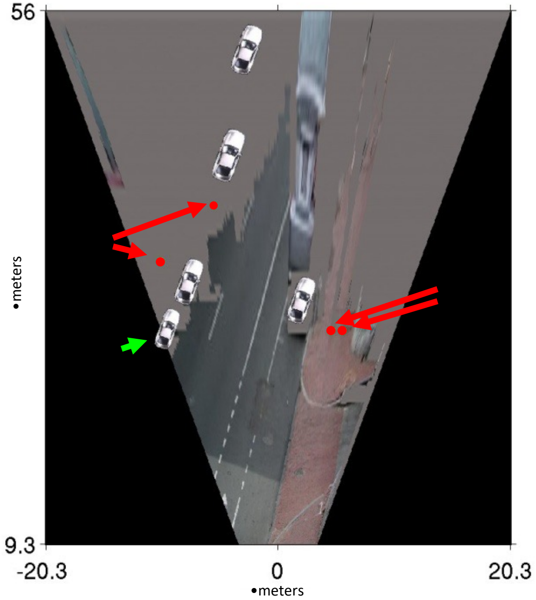
•Health services

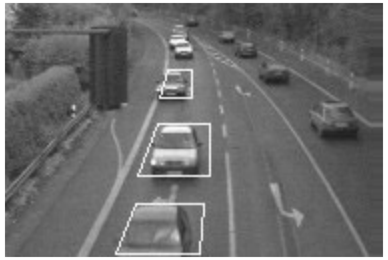


•Tourism

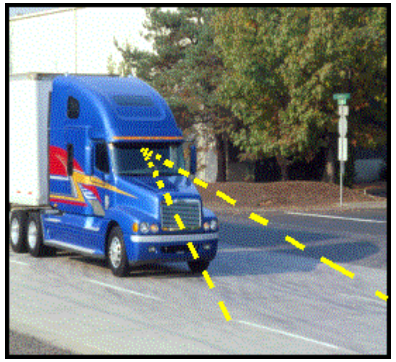
ASSISTIVE DRIVING

- Pedestrian and car detection





- Lane detection

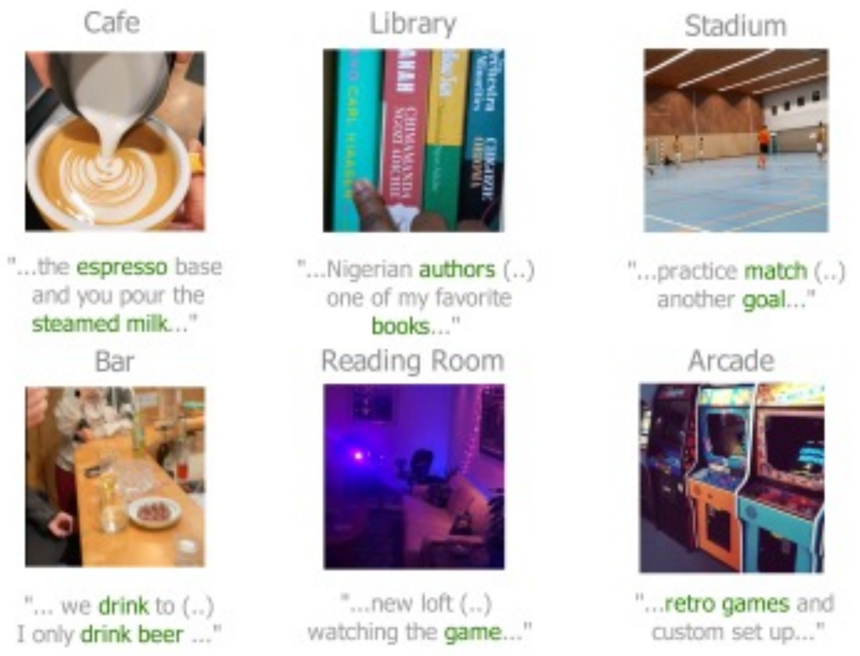


- Collision warning systems with adaptive cruise control,
- Lane departure warning systems,
- Rear object detection systems,

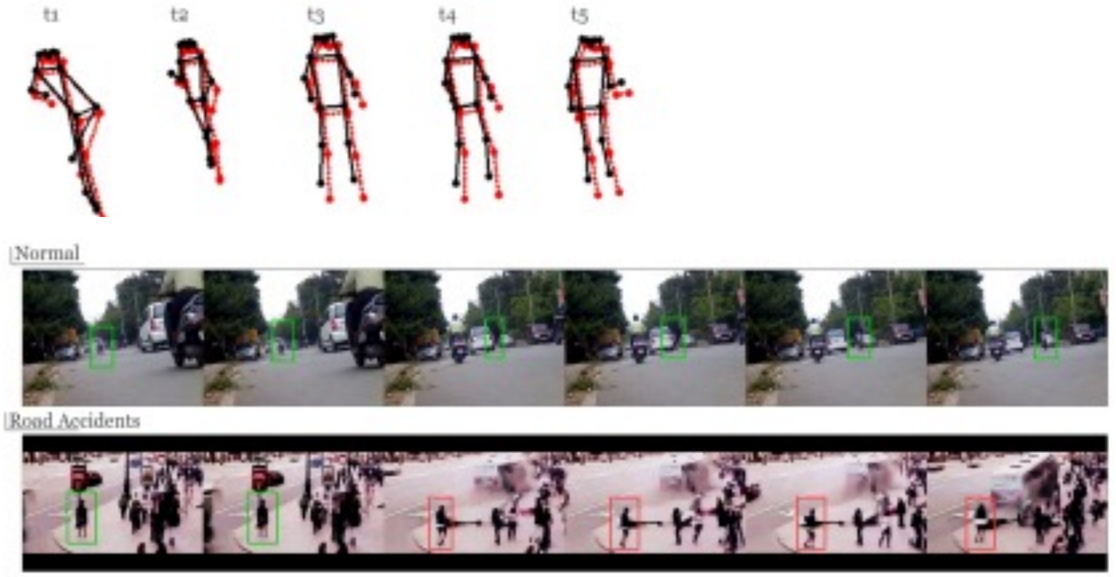
•Slide credit Fei-Fei, Fergus, Torralba CVPR07 Short Course

My work

Recorded Egocentric Photostreams
Day 1
Day 2
Day 3
Day 4
Day 5
Day ...



Cafe
"...the espresso base and you pour the steamed milk..."
Library
"...Nigerian authors (..) one of my favorite books..."
Stadium
"...practice match (..) another goal..."
Bar
"... we drink to (..) I only drink beer ..."
Reading Room
"...new loft (..) watching the game..."
Arcade
"...retro games and custom set up..."

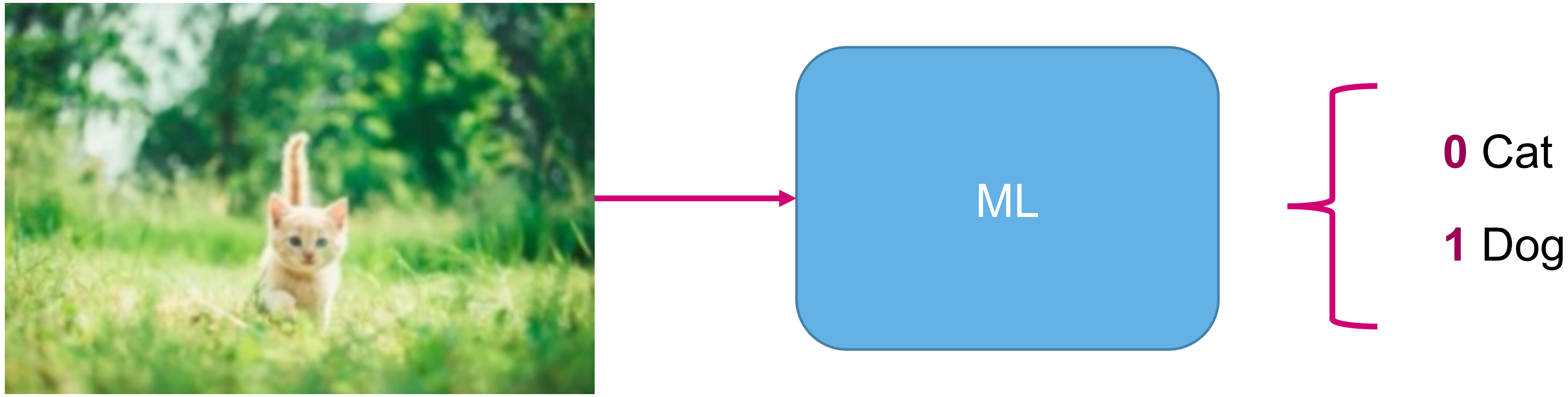


t1
t2
t3
t4
t5
Normal
Road Accidents

2 IMAGE CLASSIFICATION

Image classification framework

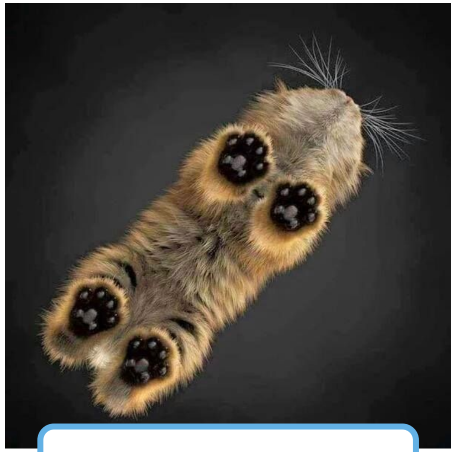
Images (+labels) + supervised ML = Image classification



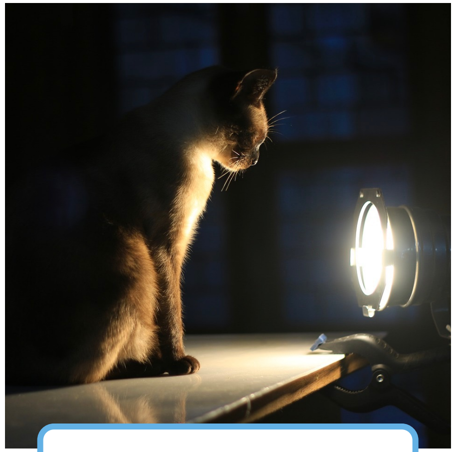
Images: Weber, Ethan, et al. "Detecting natural disasters, damage, and incidents in the wild." *European Conference on Computer Vision*. Springer, Cham, 2020.

Variation Makes Recognition Hard?

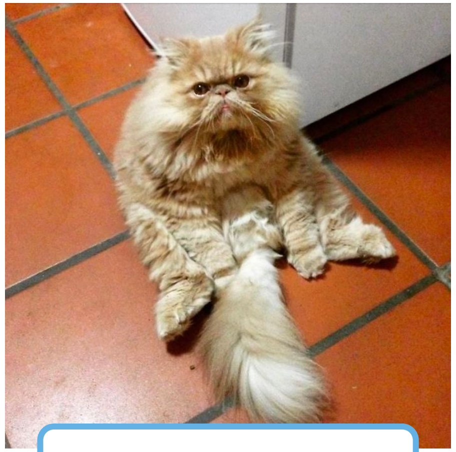
The same class of object can appear *very* differently in different images



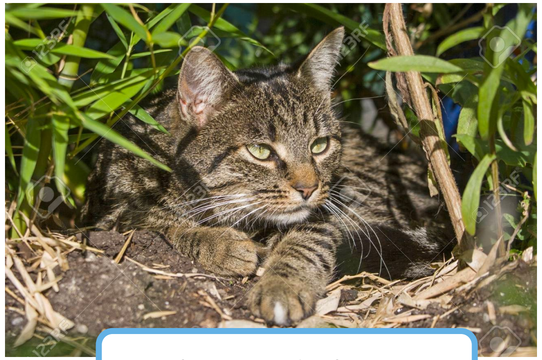
Viewpoint Variation



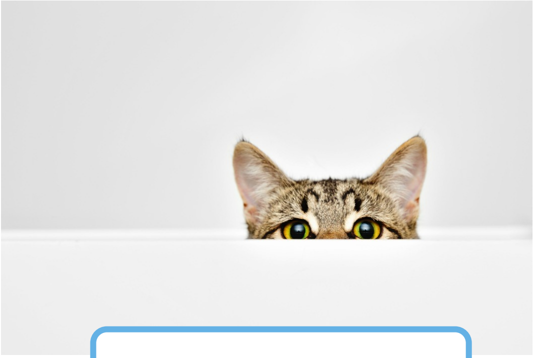
Lighting Variation



Deformation



Background Clutter

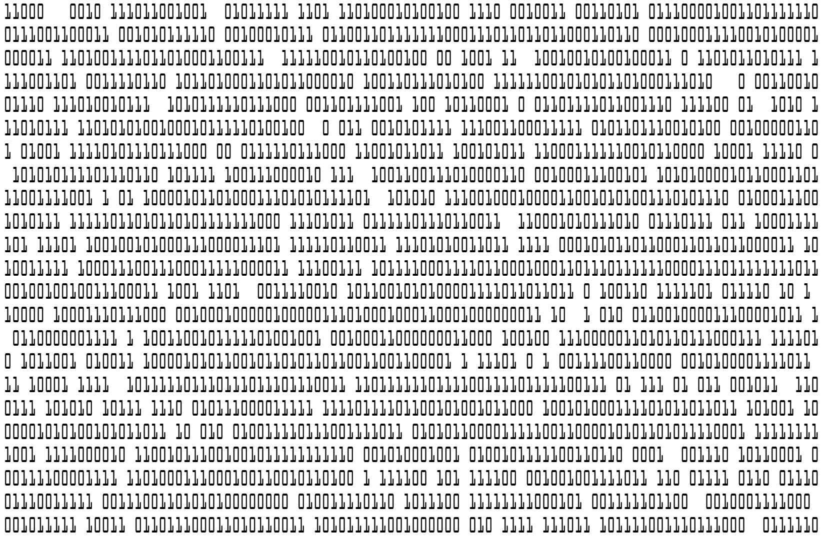


Occlusion

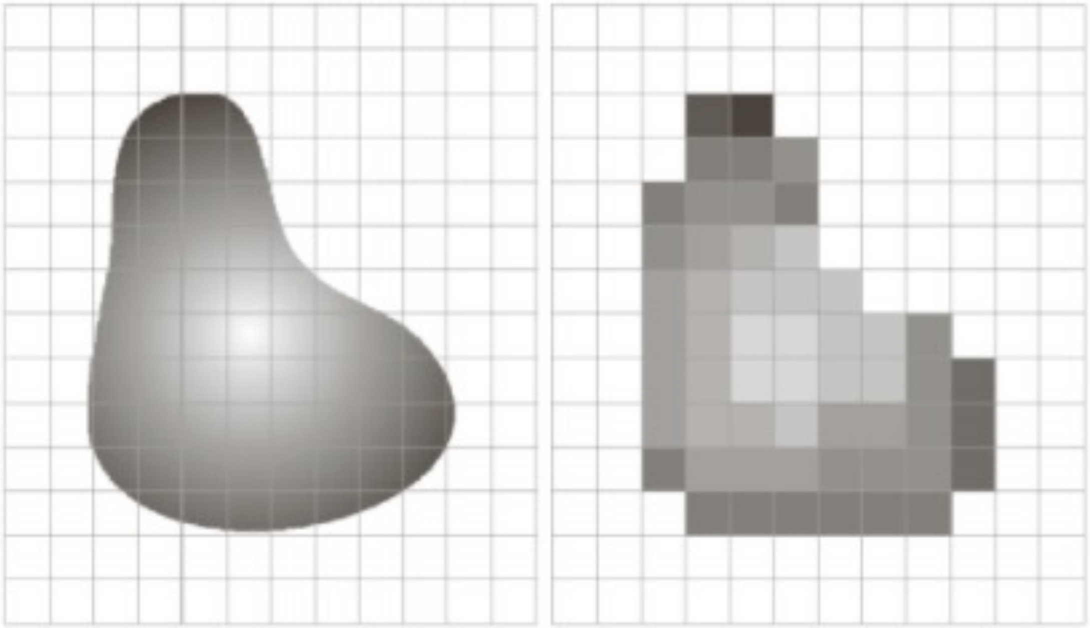
The semantic gap

What we see

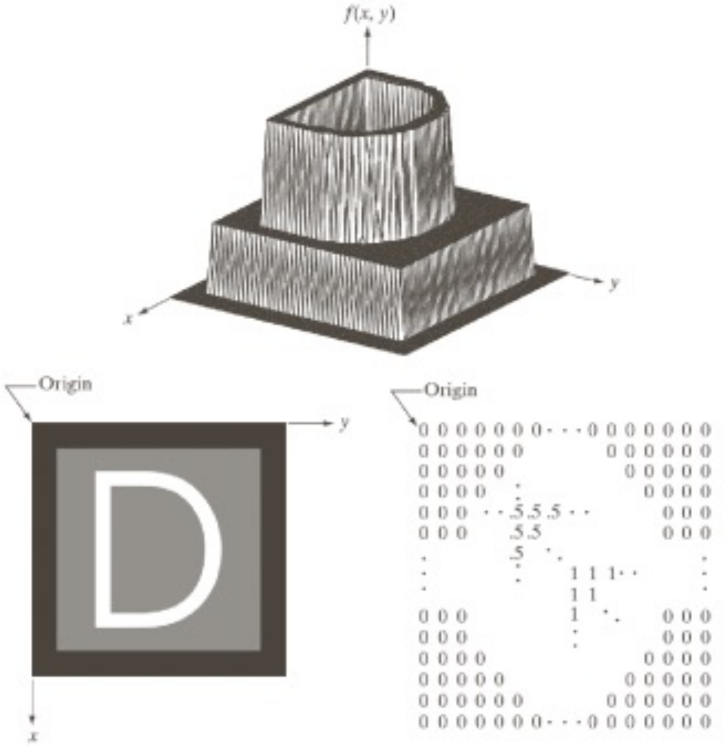
What the computer sees



Digital image and image representations



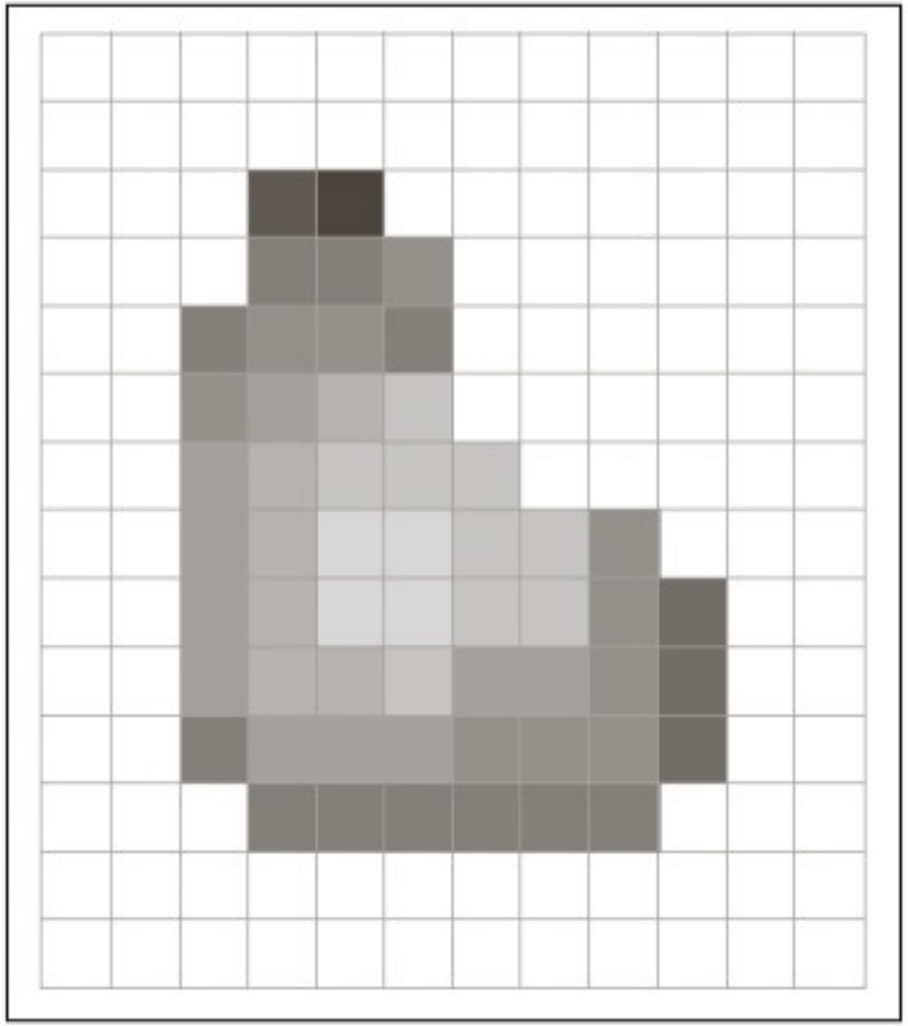
(a) Continuous image projected on sensor array. (b) Result of sampling & quantization.



(a) Surface. (b) Visual intensity array. (c) Numerical 2D array.

Images from the book Digial Image Processing 4ed by Rafael C. Gonzalez and Richard E. Woods

Digital image



Discrete domain, discrete range.

An image is a spatial representation of an object or a two- or three-dimensional scene.

The points (x, y) are called pixels, the values f(x, y) are called grey levels: for 8-bit images these grey levels range from 0 to 255.

Images from the book Digial Image Processing 4ed by Rafael C. Gonzalez and Richard E. Woods

Image representations

Image with M rows and N columns:

$$\mathbf{f} = \begin{pmatrix} f(0,0) & f(0,1) & \dots & f(0,N-1) \\ f(1,0) & f(1,1) & \dots & f(1,N-1) \\ \vdots & \vdots & \vdots & \vdots \\ f(M-1,0) & f(M-1,1) & \dots & f(M-1,N-1) \end{pmatrix}$$

Element $f(i,j)$ is value of pixel on row i and column j.

1	4	7
2	5	8
3	6	9

3×3 image

$$\mathbf{f} = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}$$

matrix repr.

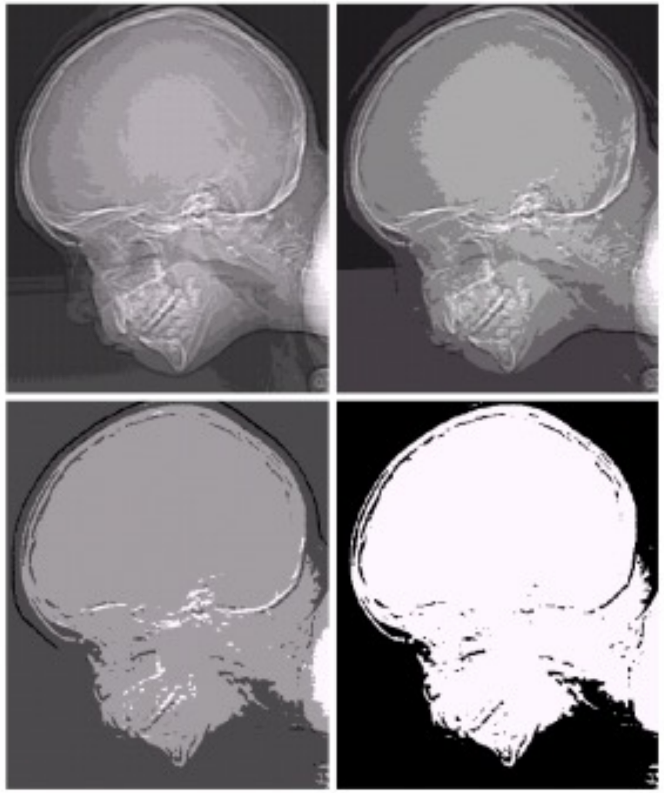
$$\vec{f} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{pmatrix}$$

vector repr.

Resolution

- **Spatial resolution:** number of pixels per unit distance (dpi: dots per inch)
- **Intensity resolution:** number of bits used to quantize the intensity range

(a) 1250 dpi. (b) 300 dpi. (c) 150 dpi. (d) 72 dpi.



(a) 16 levels. (b) 8 levels. (c) 4 levels. (d) 2 levels.

Images from the book Digial Image Processing 4ed by Rafael C. Gonzalez and Richard E. Woods

Color image

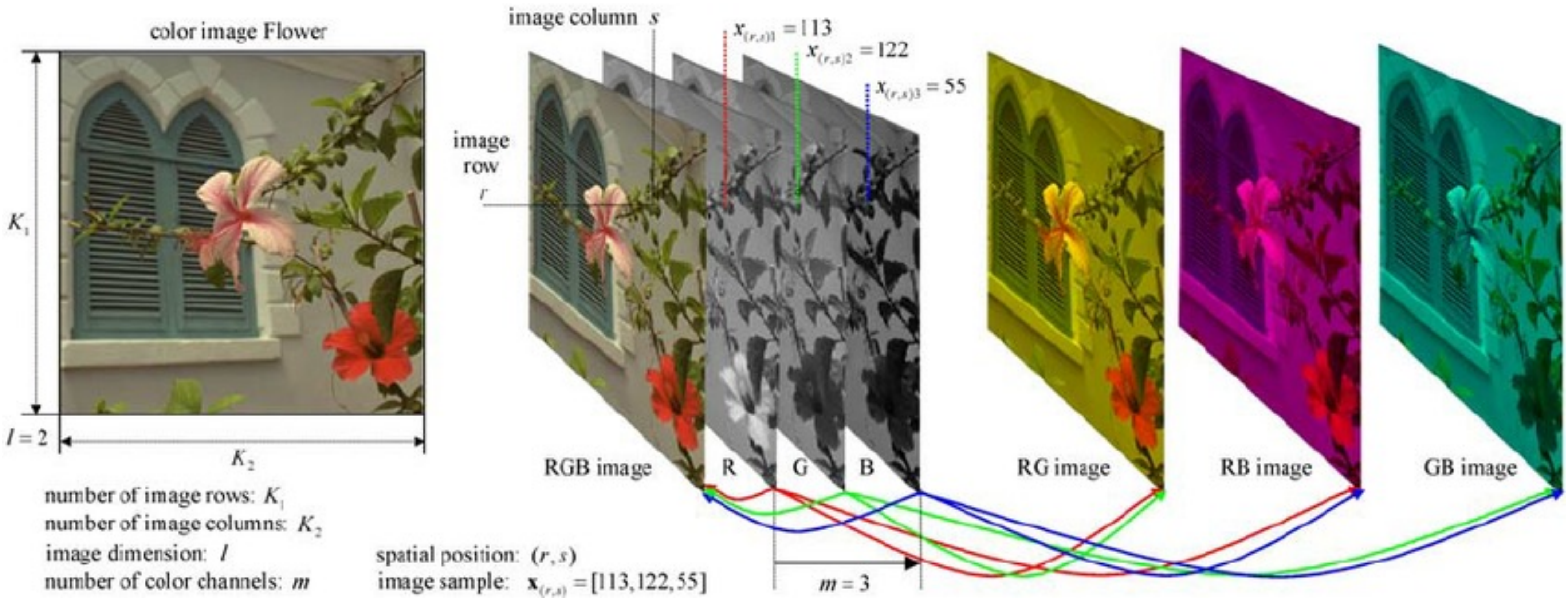


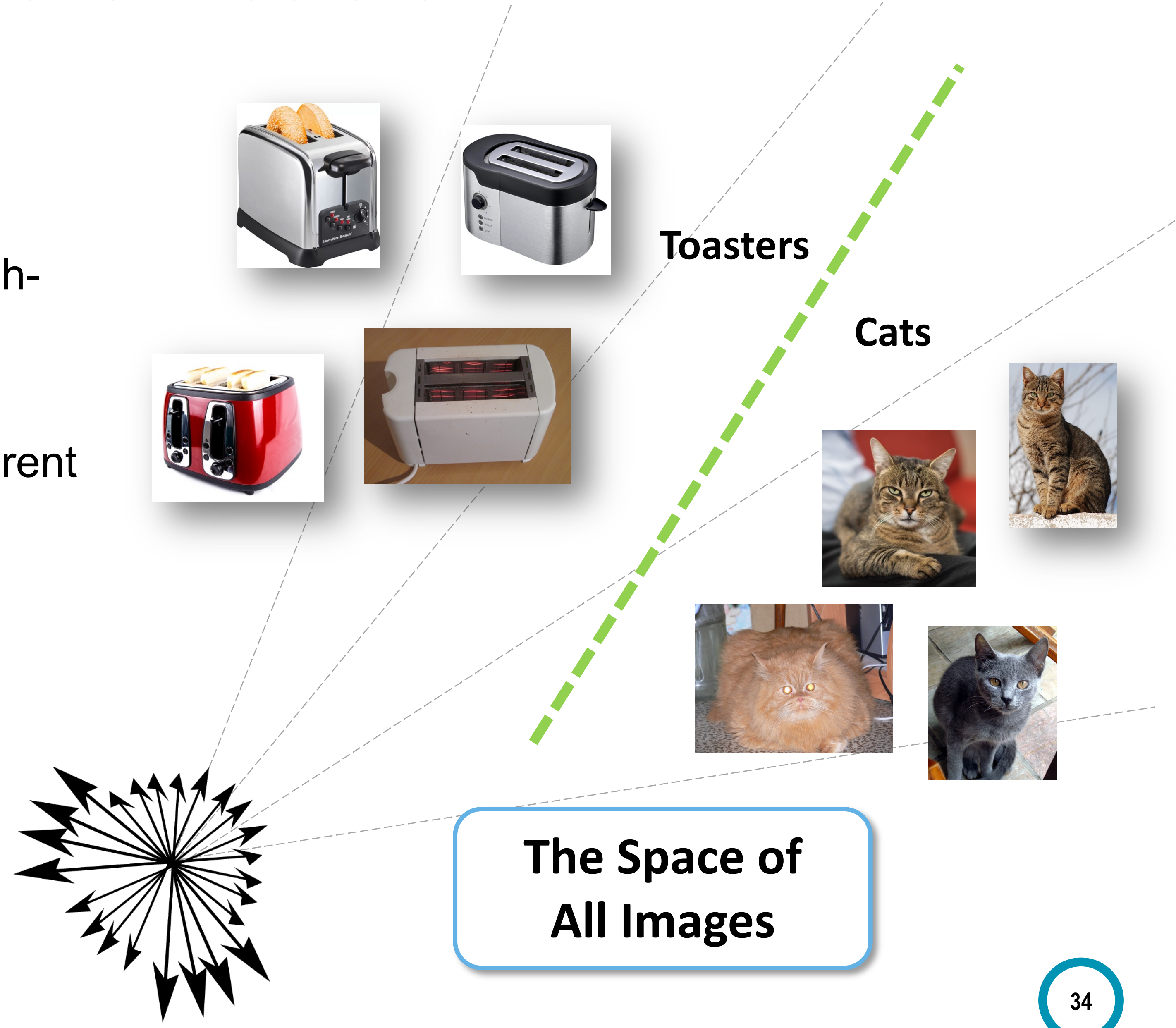
Image from Lukac, Rastislav, and Konstantinos N. Plataniotis. "Single-sensor camera image processing." *Color Image Processing*. CRC Press, 2018. 409-438.

Images As High-Dimensional Vectors

- An image is just a bunch of numbers
- Let’s stack them up into a vector
  - Our training data is just a bunch of high-dimensional points now

Images As High-Dimensional Vectors

- An image is just a bunch of numbers
- Let's stack them up into a vector
  - Our training data is just a bunch of high-dimensional points now
- Divide space into different regions for different classes



Images As High-Dimensional Vectors

- An image is just a bunch of numbers
- Let's stack them up into a vector
    - Our training data is just a bunch of high-dimensional points now

- Divide space into different regions for different classes

Images As High-Dimensional Vectors

- An image is just a bunch of numbers
- Let's stack them up into a vector
  - Our training data is just a bunch of high-dimensional points now
- Divide space into different regions for different classes

or

- Define a distribution over space for each class

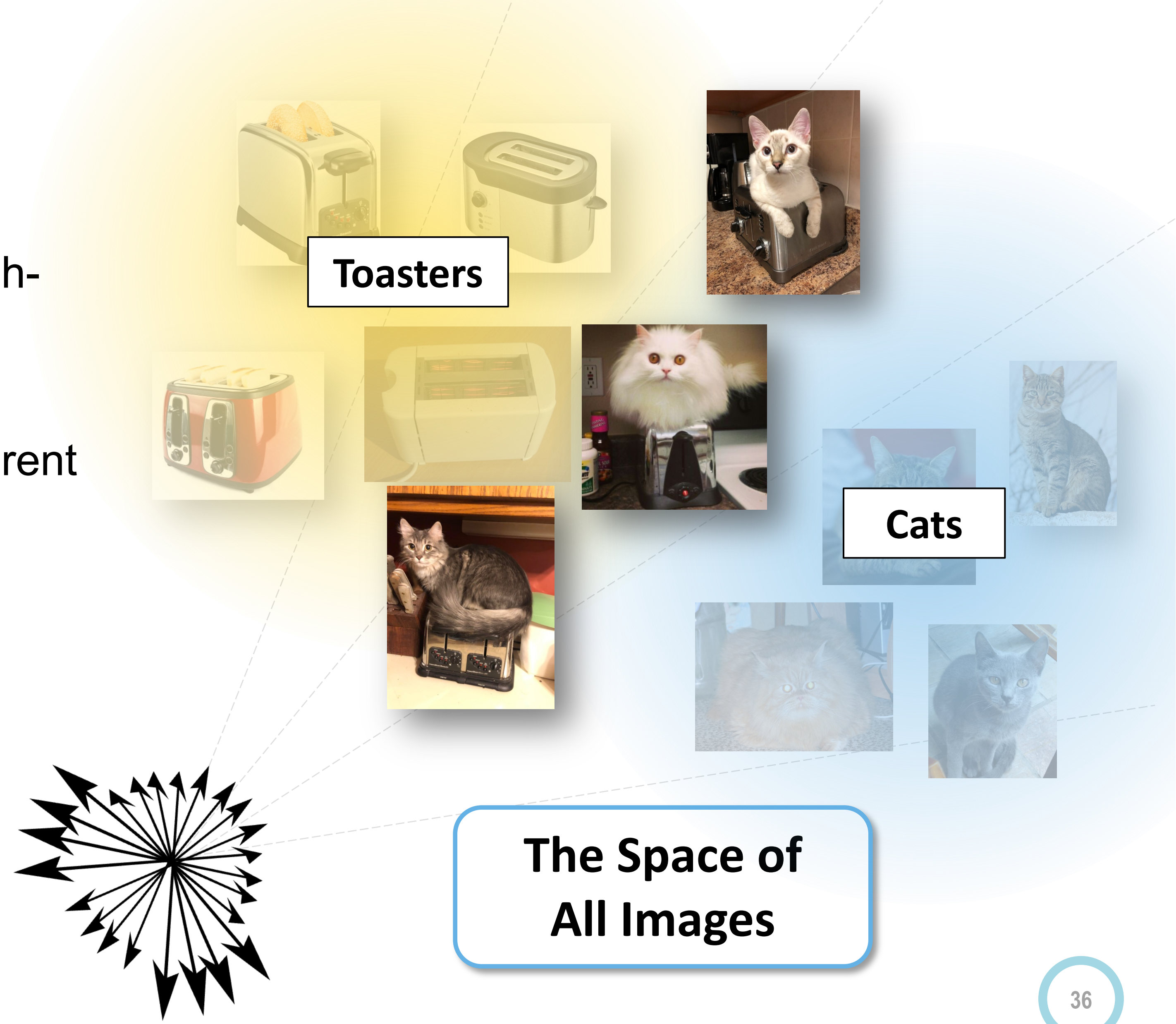
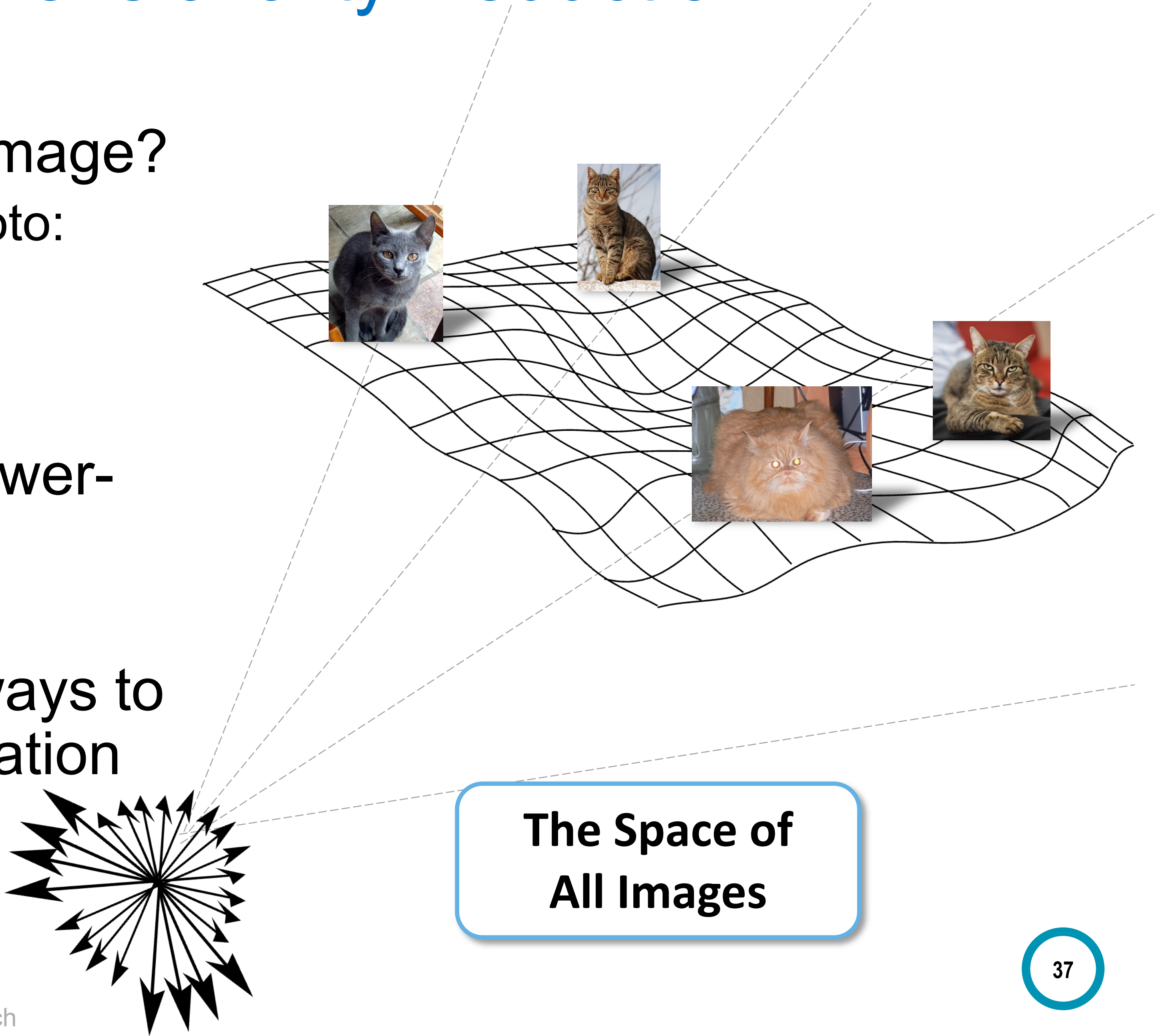
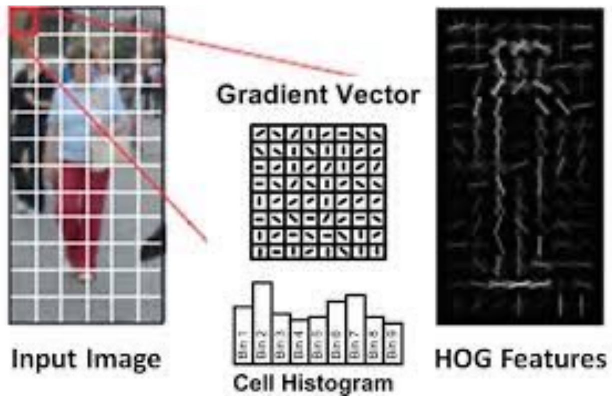


Image Features and Dimensionality Reduction

- How high-dimensional is an image?
  - Let's consider an iPhone X photo:
    - 4032 x 3024 pixels
    - Every pixel has 3 colors
    - 36,578,304 pixels (36.5 Mega pixels)
- In practice, images sit on a lower-dimensional manifold
- Think of image features and dimensionality reduction as ways to represent images by their location on such manifolds

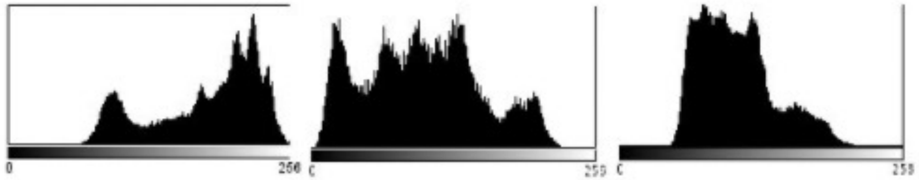


Feature extraction

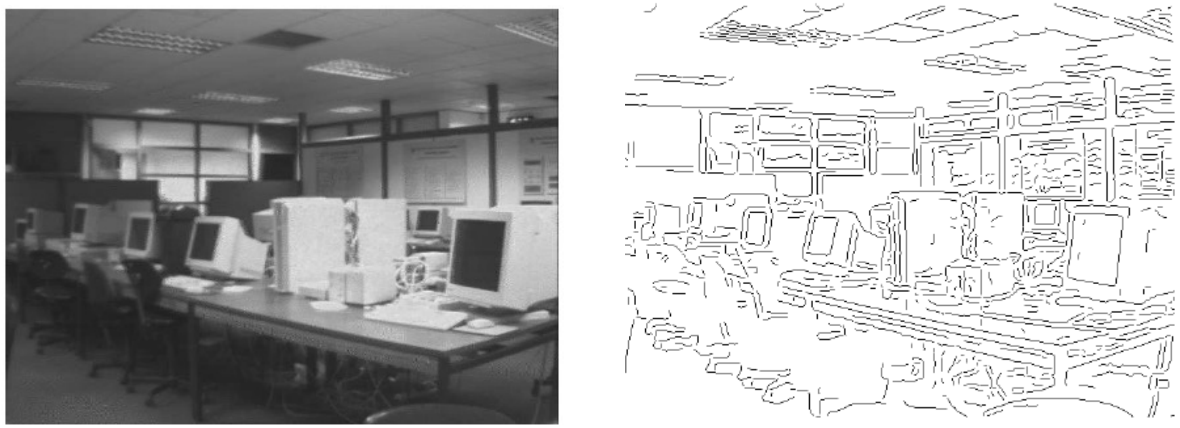


Histogram of oriented gradients
https://arxiv.org/pdf/1703.05853.pdf

$$F_R = [f_{R0}, \ldots, f_{R255}]$$
$$F_G = [f_{G0}, \ldots, f_{G255}]$$
$$F_B = [f_{B0}, \ldots, f_{B255}]$$
$$F_{RGB} = [F_R \mid F_G \mid F_B]$$



RGB histograms as descriptor



Edges and corners

-1	-2	-1
0	0	0
+1	+2	+1

-1	0	+1
-2	0	+2
-1	0	+1

Kernels for the Sobel operator.

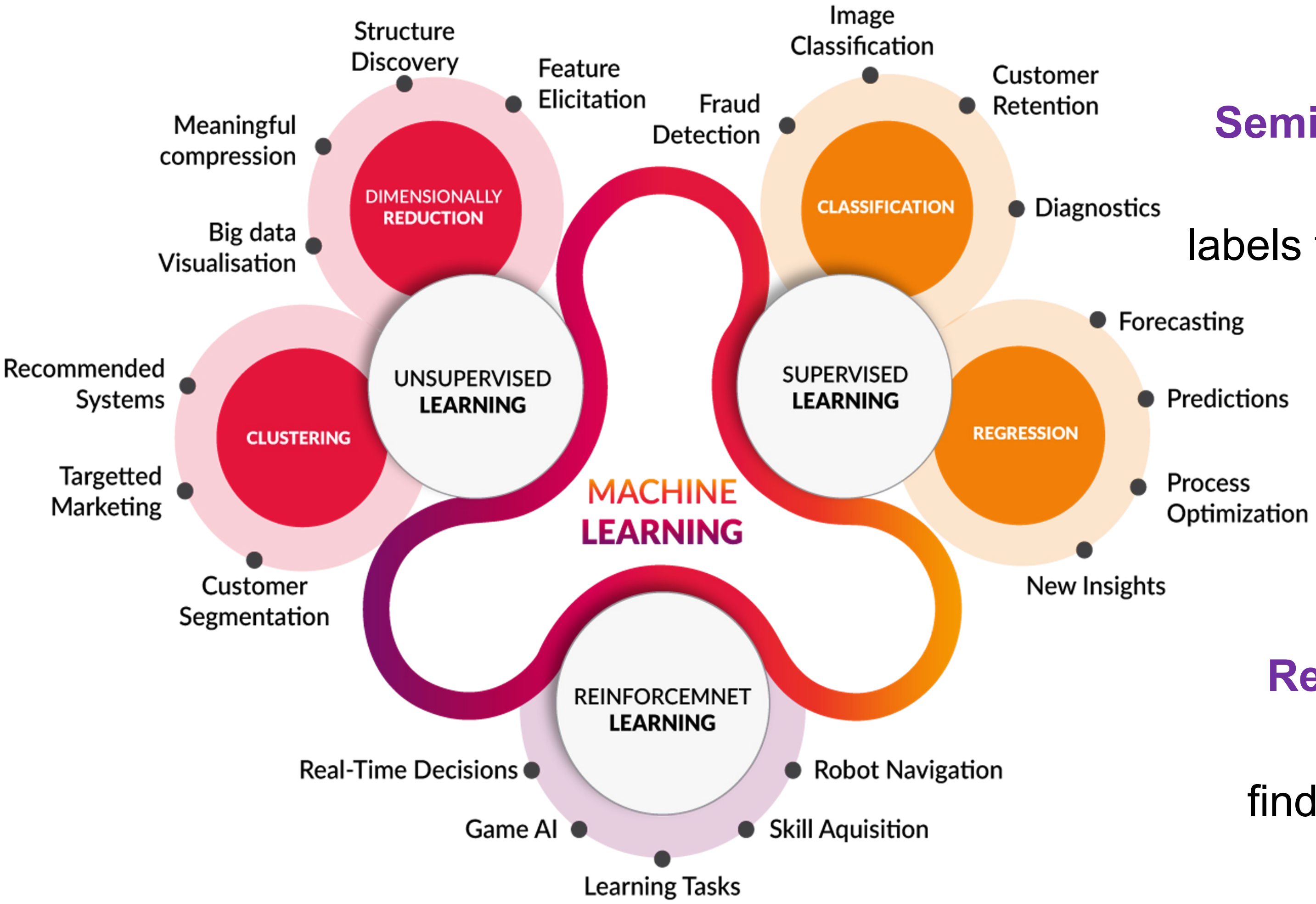
3 MACHINE LEARNING

Types of machine learning problems

Unsupervised learning

input data, no labels

- Clustering: group similar data points.
- Density estimation
- Dimensionality reduction
- Outlier/novelty detection



Supervised learning

labels are provided

- Classification
- Regression

Semi-Supervised learning

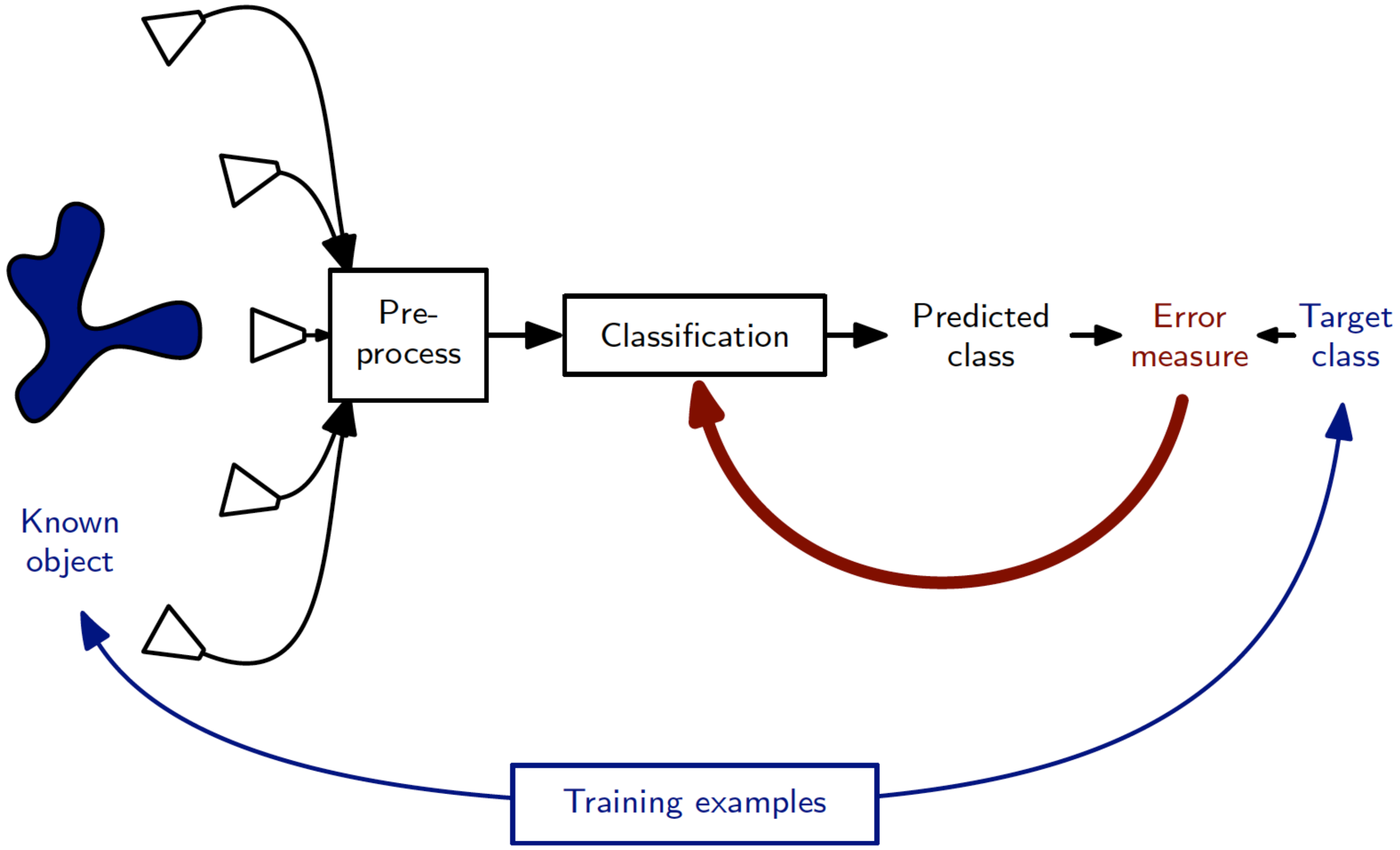
labels for just part of the data

Reinforcement learning

find a sequence of actions (policy) that reaches a ta .

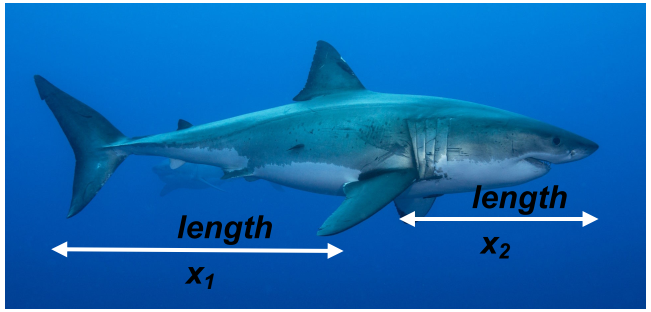
*Image from https://towardsdatascience.com/coding-deep-learning-for-beginners-types-of-machine-learning-b9e651e1ed9d

Supervised Training - Classification



Supervised learning

data: observations, e.g.
vectors of num. values



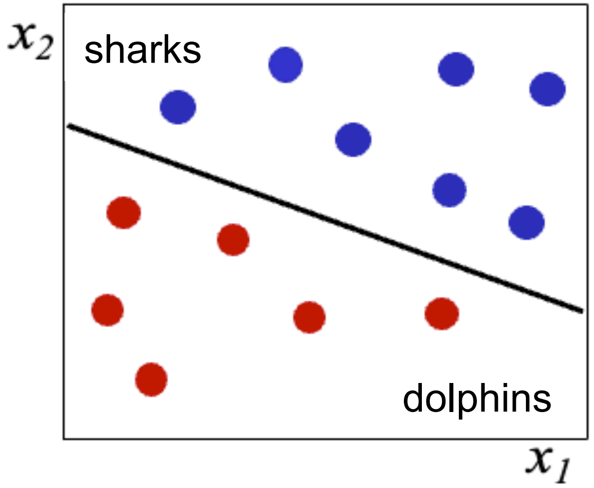
example data set

classification tasks:
assign data to a **category**

model: linear separation

$$\alpha x_1 + \beta x_2 > \gamma \text{ “shark”}$$

else “dolphin”

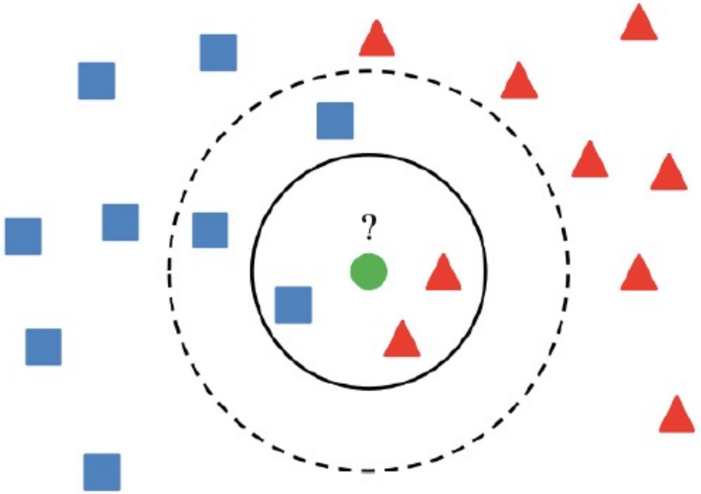


training:

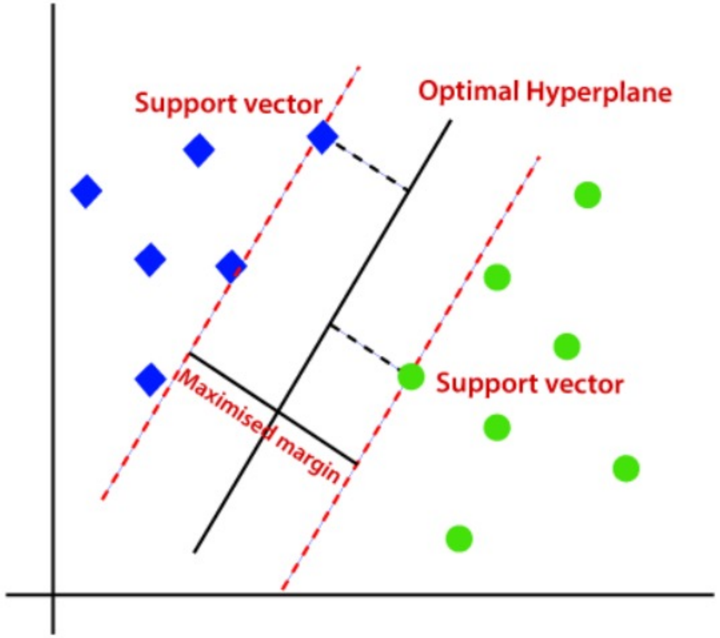
⇨ optimize parameters α, β, γ

Supervised learning methods for classification

K-NN



SVM



Scikit-learn python packages - https://scikit-learn.org/stable/supervised_learning.html
https://vitalflux.com/k-nearest-neighbors-explained-with-python-examples/
https://medium.com/@viveksalunkhe80/support-vector-machine-svm-88f360ff5f38

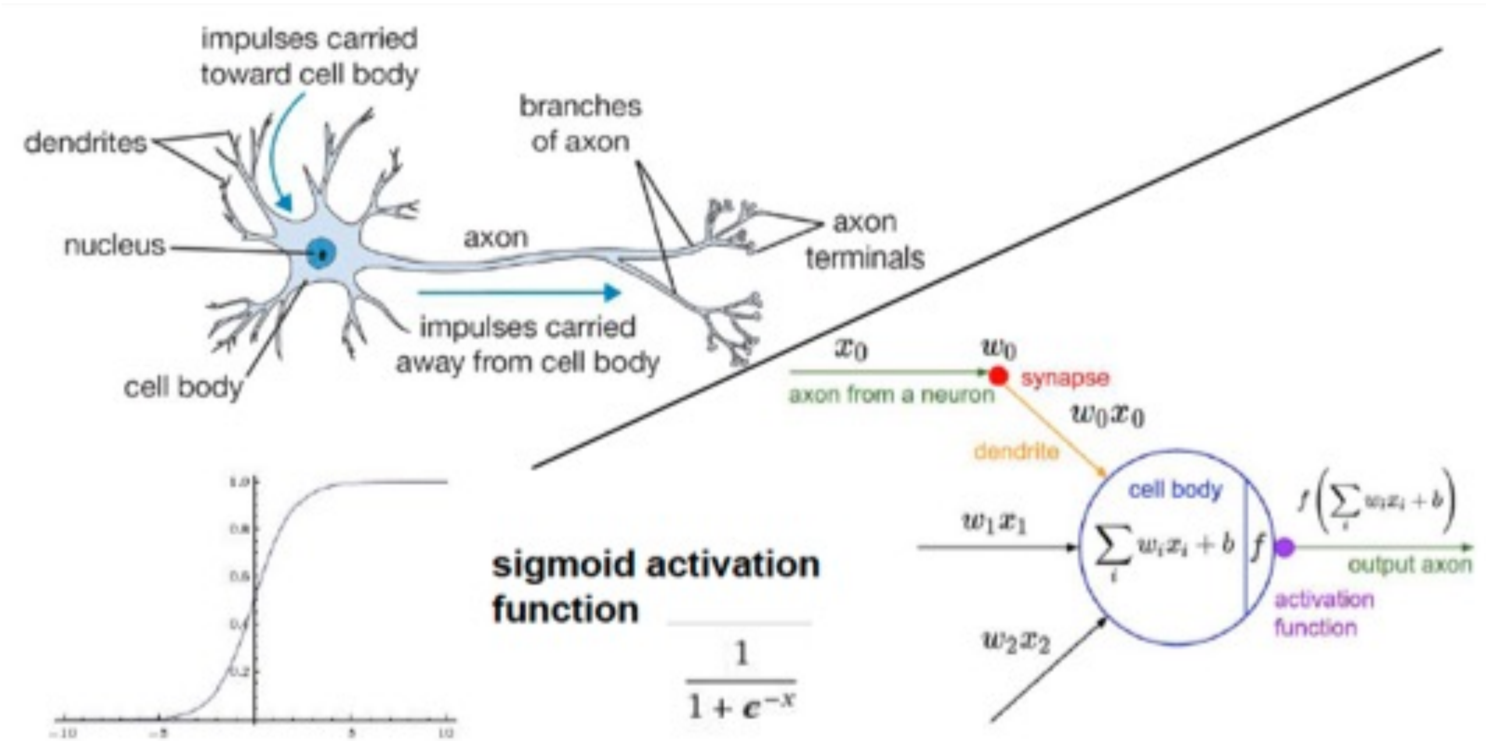
(Deep) Neural Networks in Image classification

- As a classifier
- As a feature extractor + classical ML method

Neural Networks

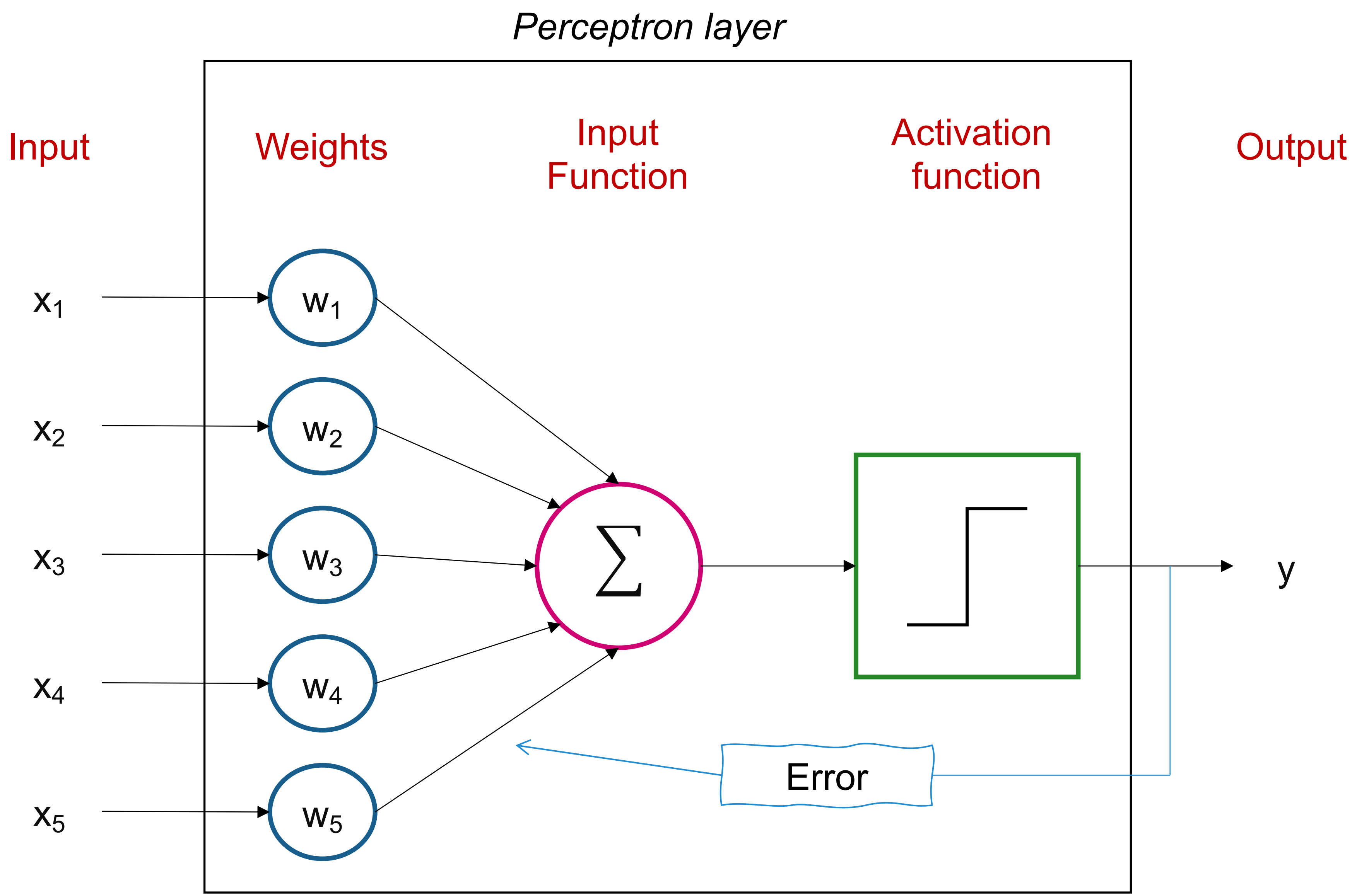
Why are they called Neural Networks?

Artificial NNs ~ 100M neurons
Human NNs ~ 5B of neurons



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

Perceptron learning algorithm



Perceptron learning algorithm

Send a signal to the output if inputs are sufficiently activated

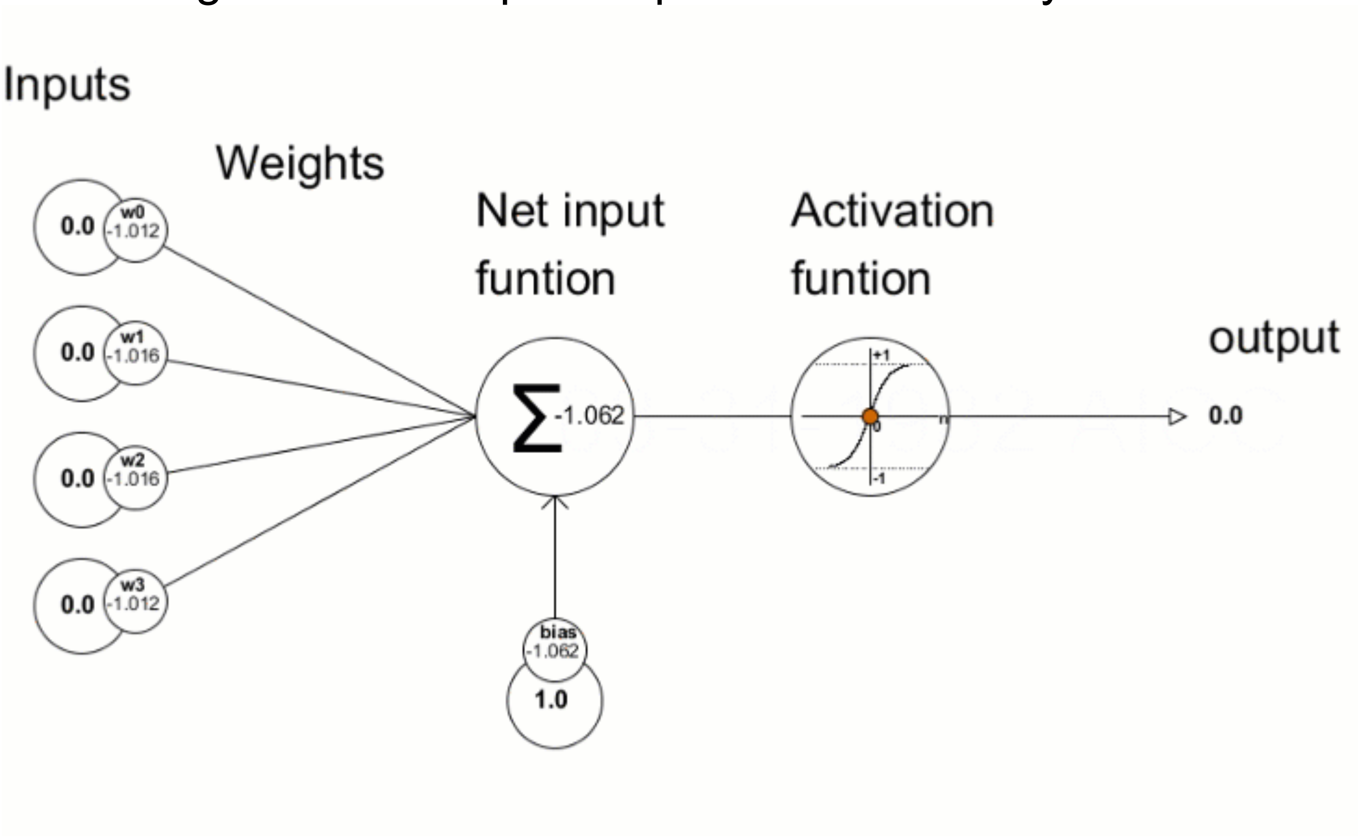
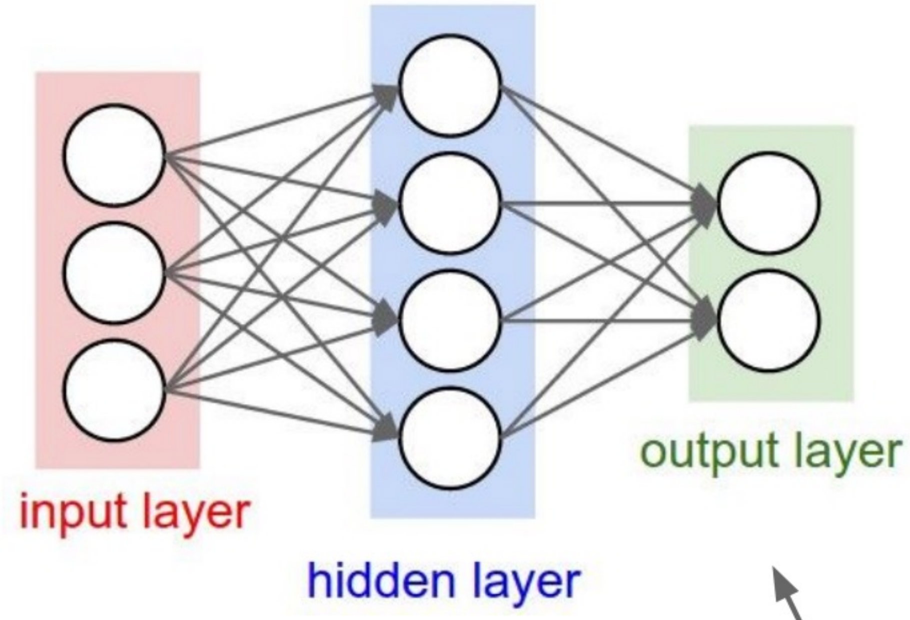


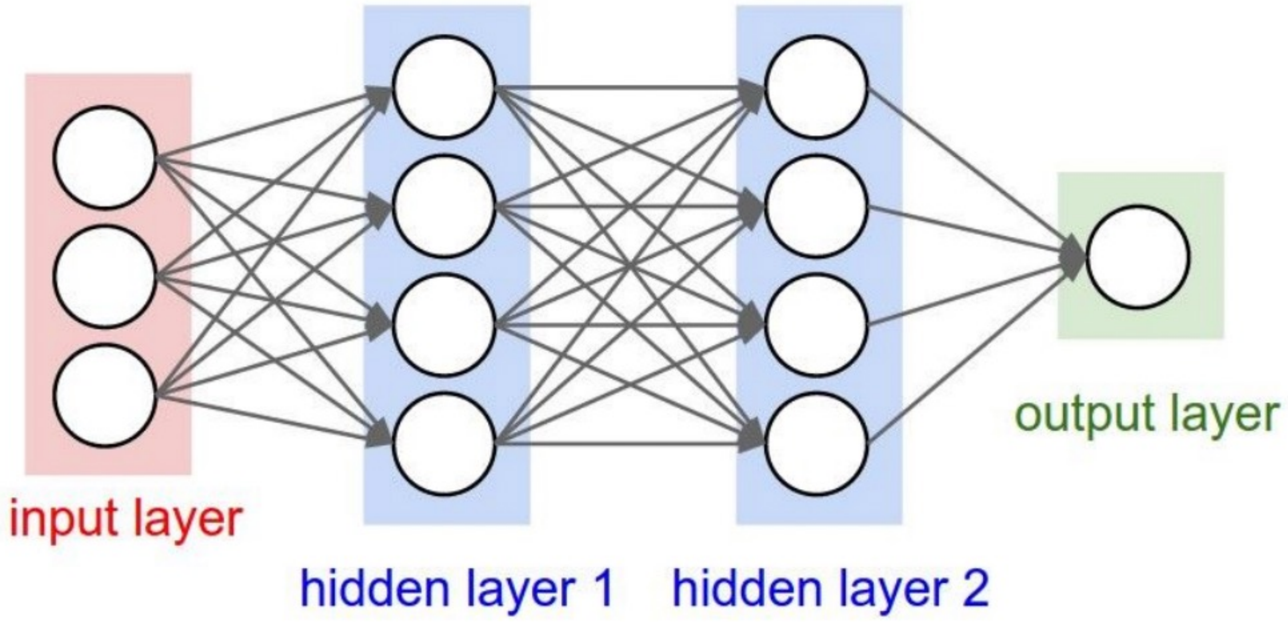
Figure: https://www.mql5.com/en/blogs/post/724245

Training data set $\{\mathbf{x}_n, t_n\}$
uses the target coding scheme:

$$t_n = \begin{cases} +1 & \text{if } \mathbf{x}_n \text{ belongs to } \mathcal{C}_1 \\ -1 & \text{otherwise} \end{cases}$$

Neural Networks Architectures



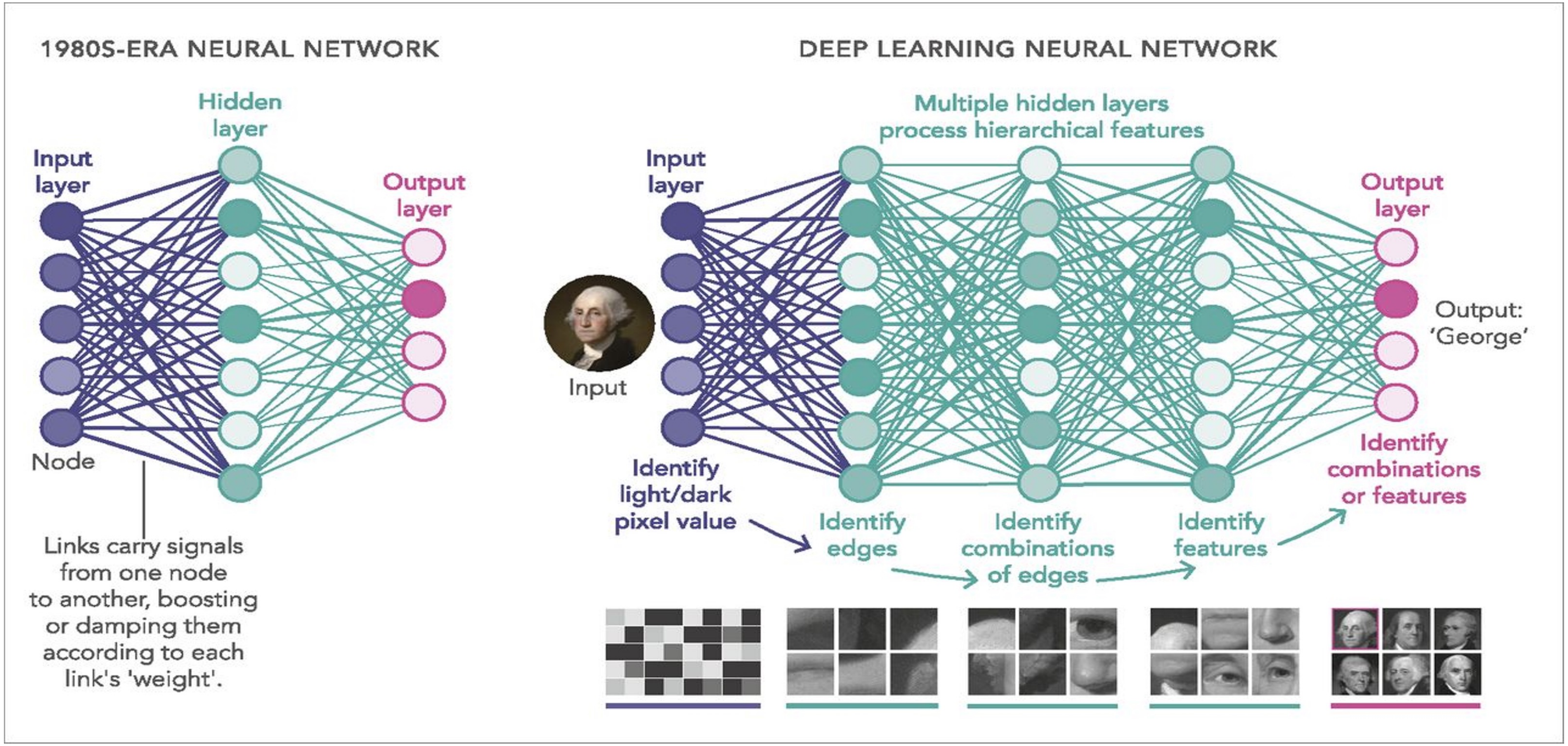


"2-layer Neural Net", or
"1-hidden-layer Neural Net"

"3-layer Neural Net", or
"2-hidden-layer Neural Net"

"Fully-connected" layers

Deep learning

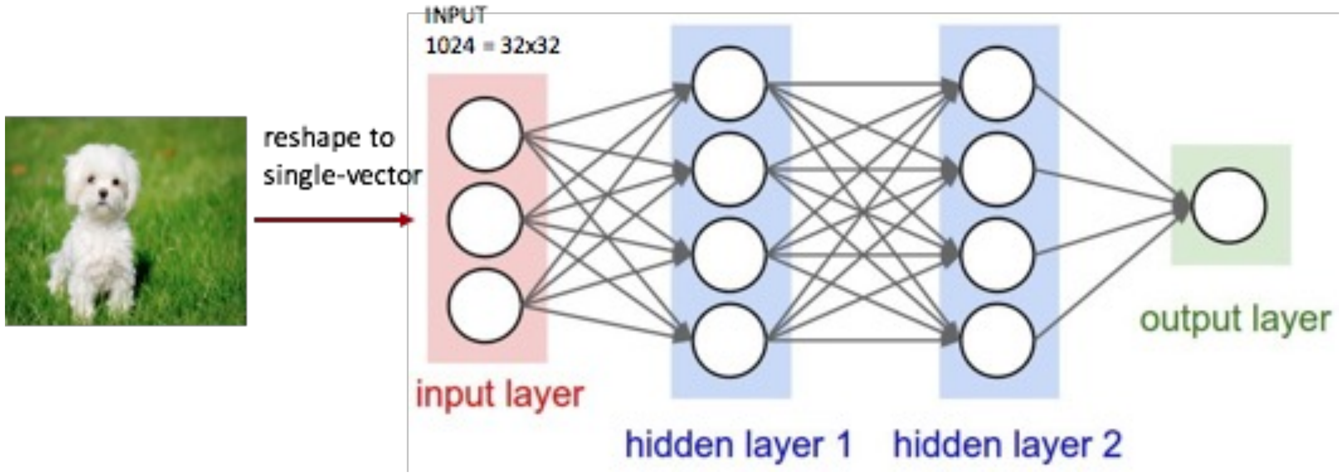


Waldrop, M. Mitchell. "News Feature: What are the limits of deep learning?." *Proceedings of the National Academy of Sciences* 116.4 (2019): 1074-1077

Artificial vs. Convolutional Neural Networks

Question : what is the difference between Convolutional Neural Networks (CNNs) and Artificial Neural Networks (NNs)?

ANN



Artificial vs. Convolutional Neural Networks

Question : what is the difference between Convolutional Neural Networks (CNNs) and Artificial Neural Networks (NNs)?

PROC. OF THE IEEE, NOVEMBER 1998

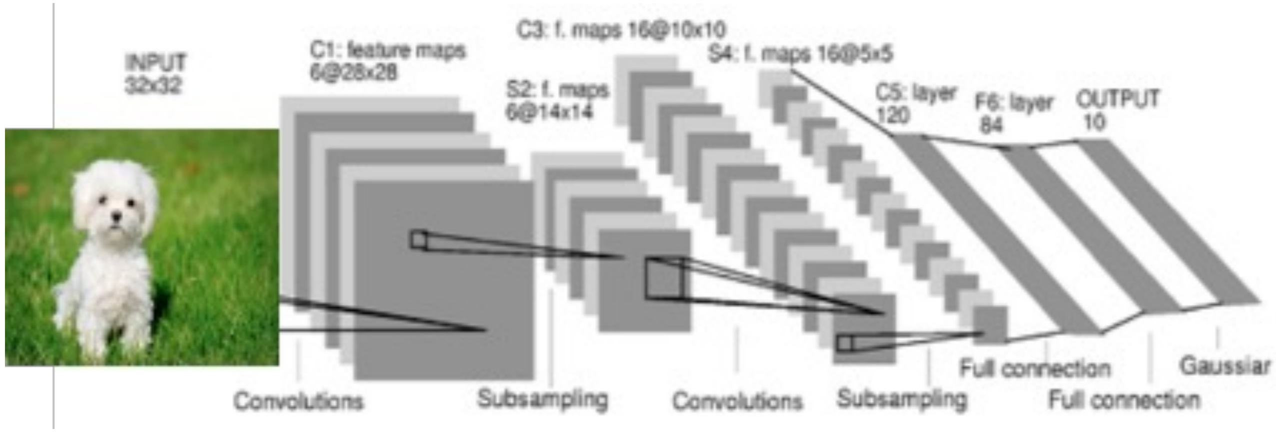
Gradient-Based Learning Applied to Document Recognition

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner

all trained ('end-to-end').

representation on different levels of detail/abstraction

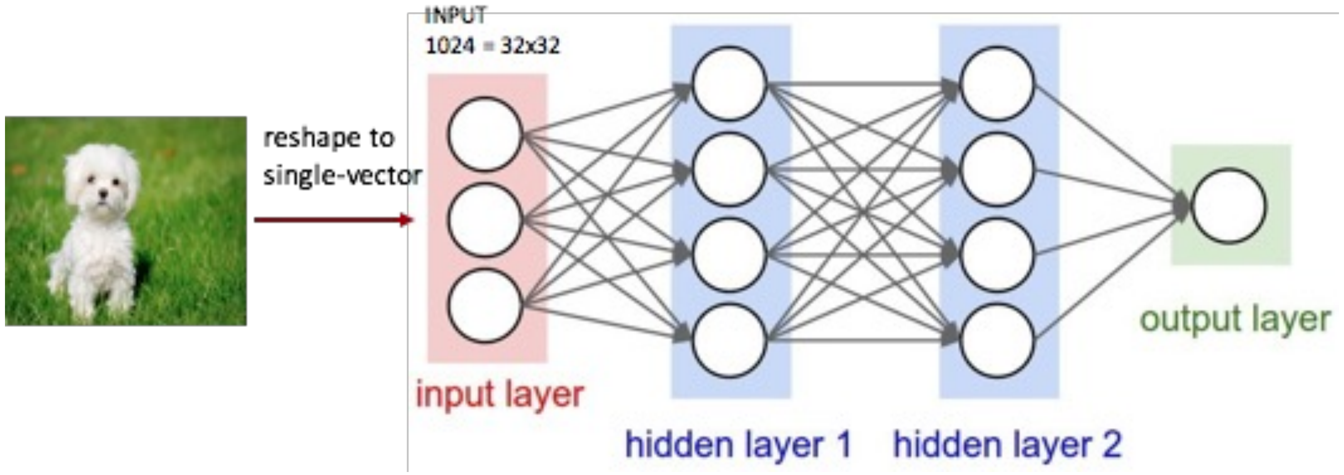
CNN



Artificial vs. Convolutional Neural Networks

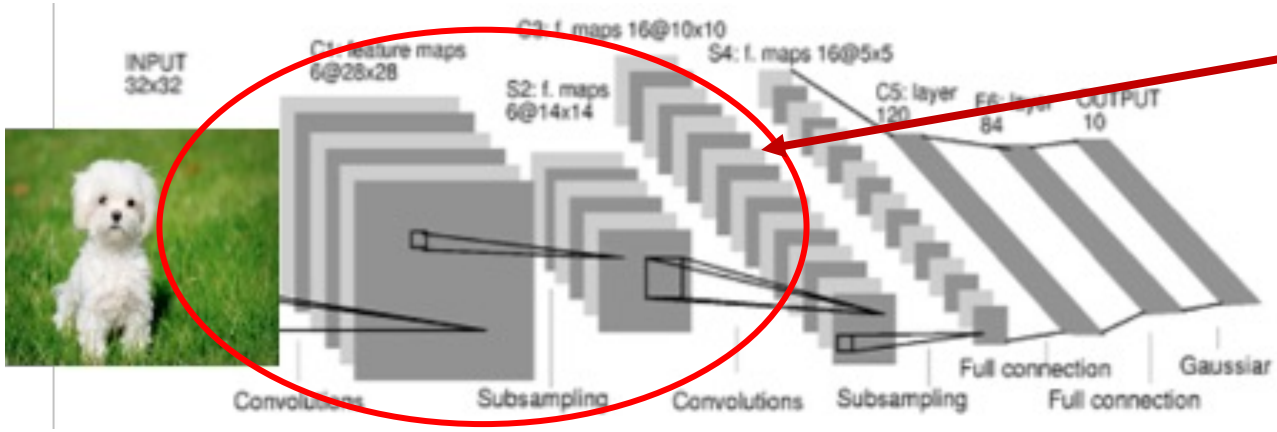
Question : what is the difference between Convolutional Neural Networks (CNNs) and Artificial Neural Networks (NNs)?

ANN



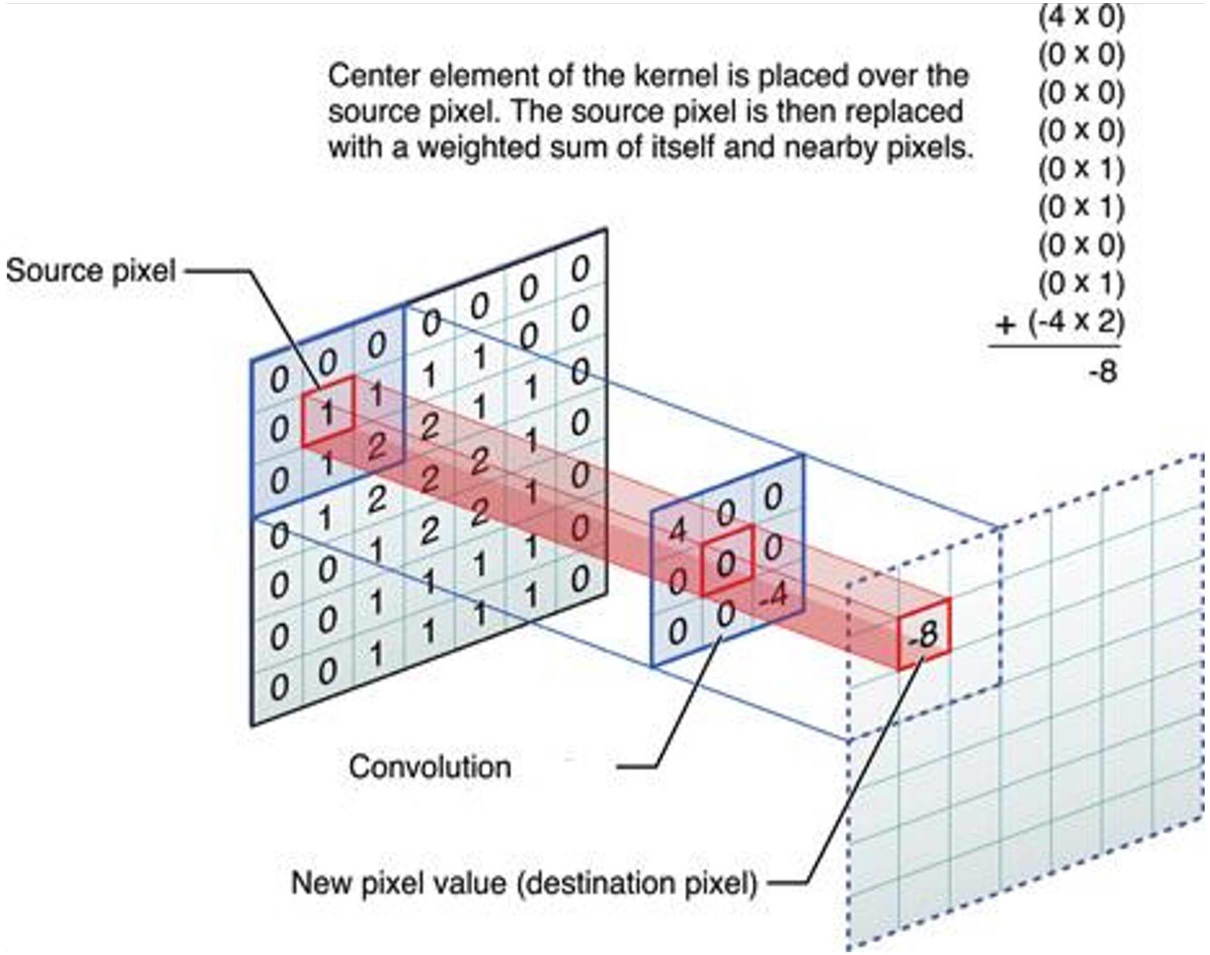
Spatial information is lost!

CNN



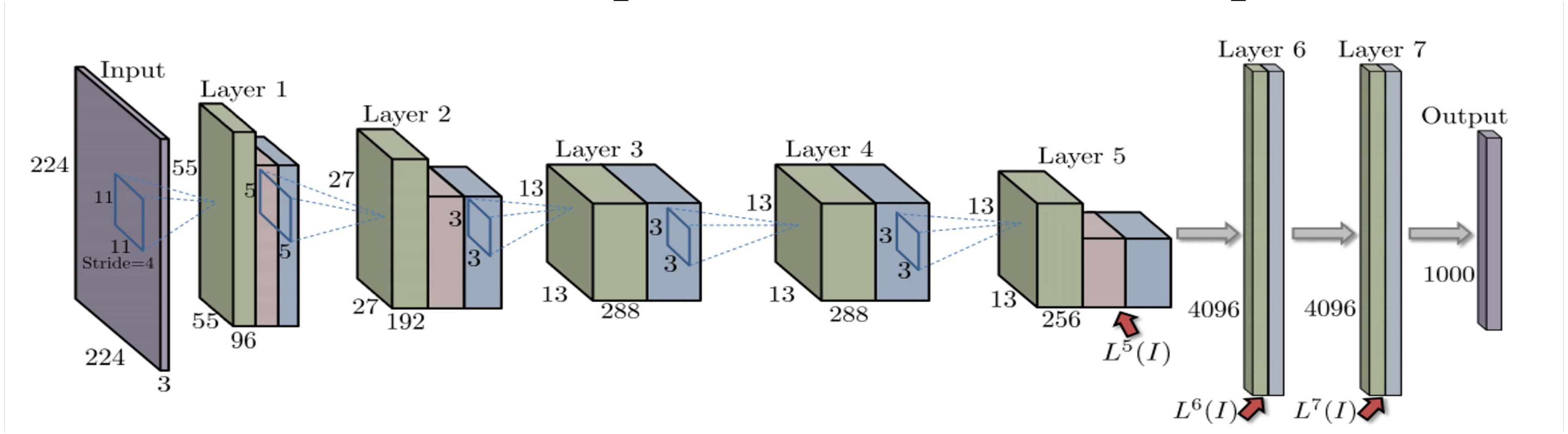
Can learn spatial correlations!

Convolutions

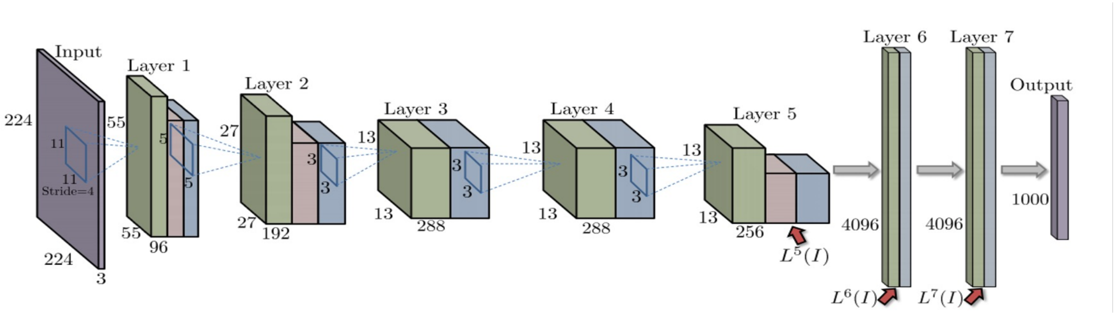


Convolutional Neural Networks

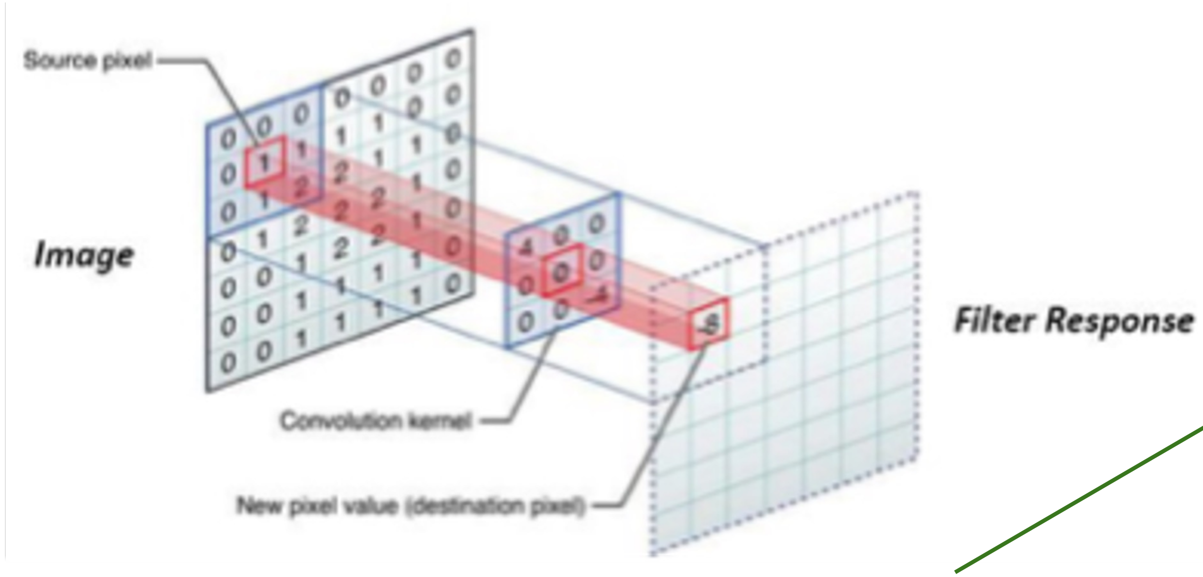
Spatial Info **No Spatial Info**

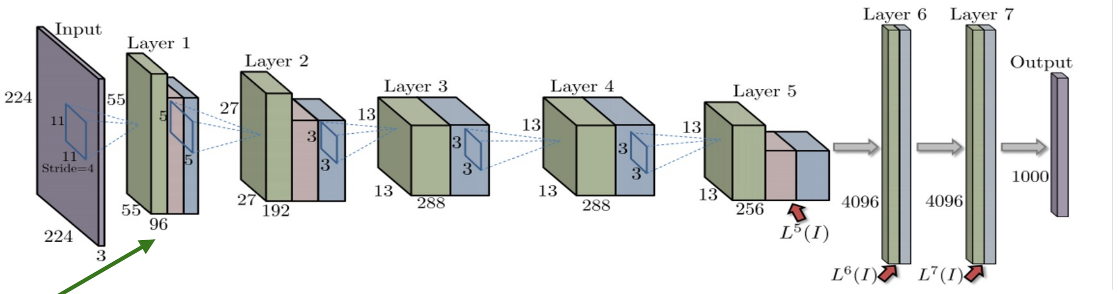


Convolutional Neural Networks



Convolutional Neural Networks

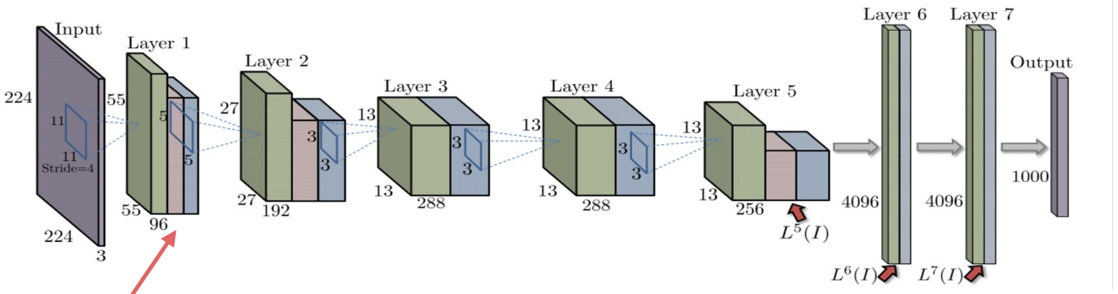




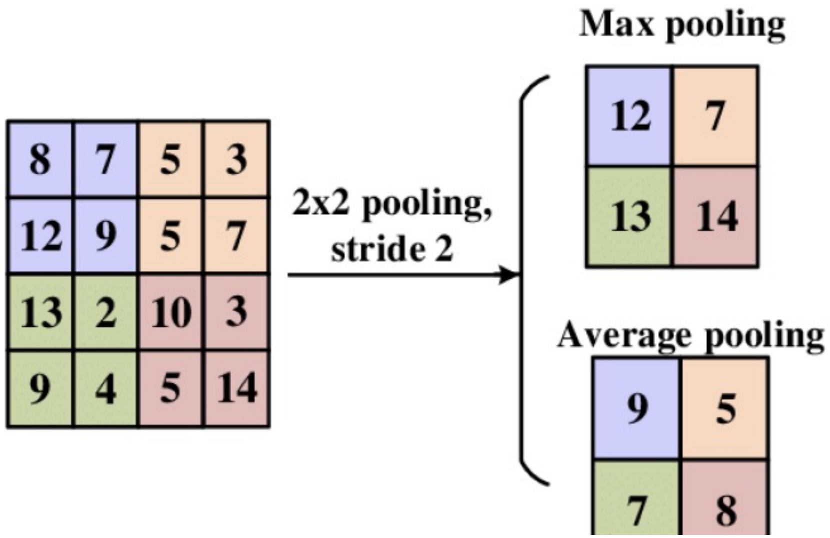
Convolutional layer

- Convolution operations with (zero-)padding and stride.

Convolutional Neural Networks

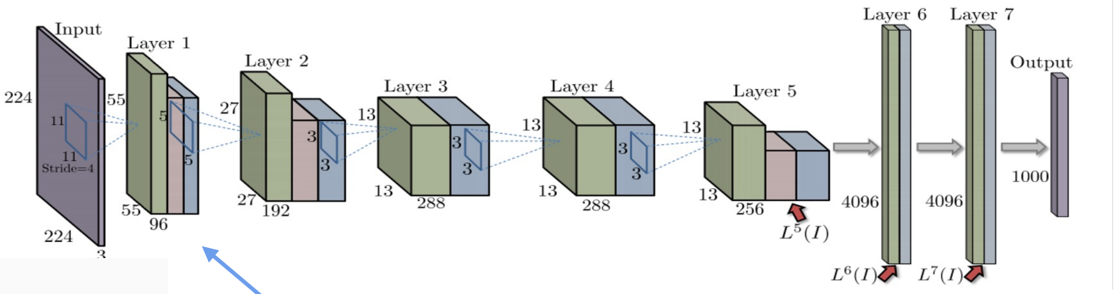


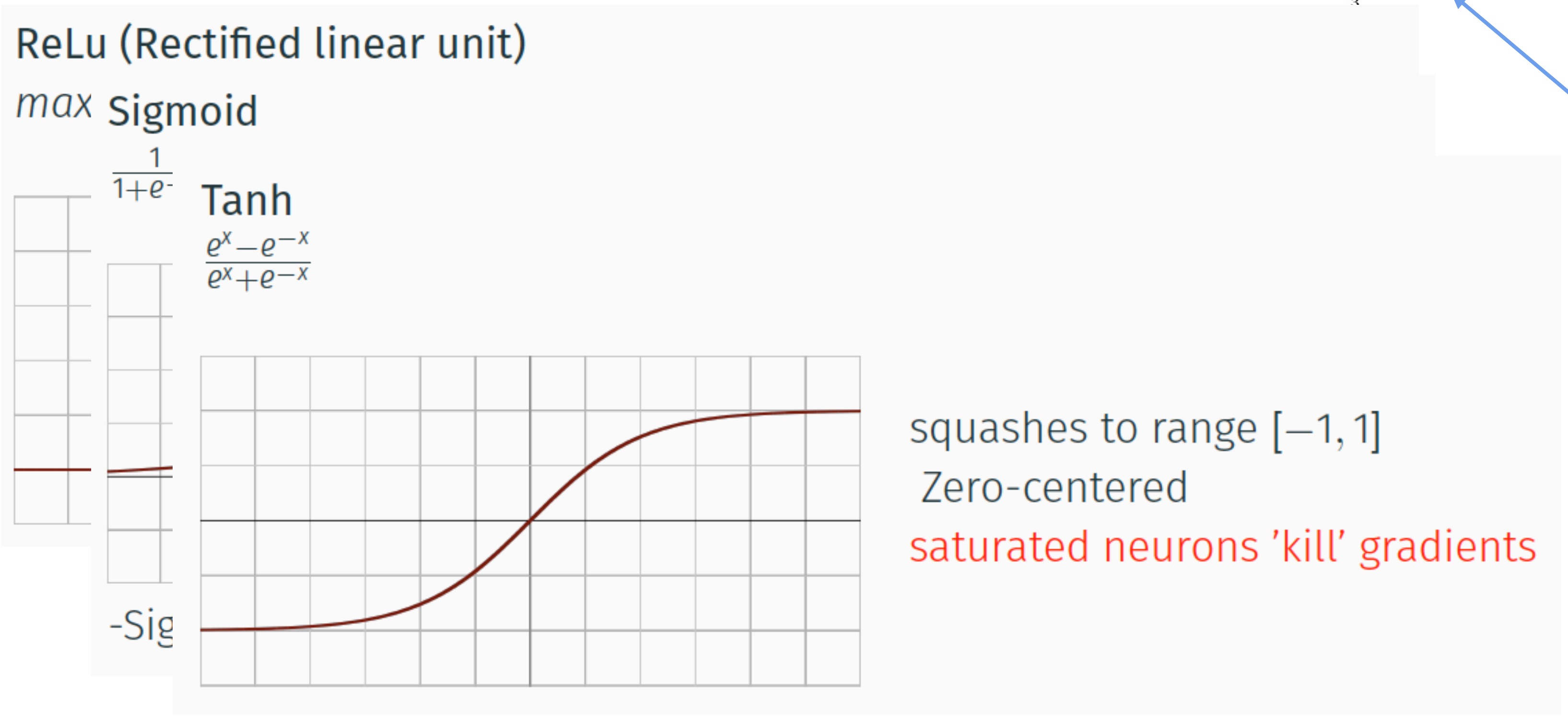
Pooling layer



- No weights to be trained
- Pooling layer needs 2 hyperparameters: kernel size F, stride S

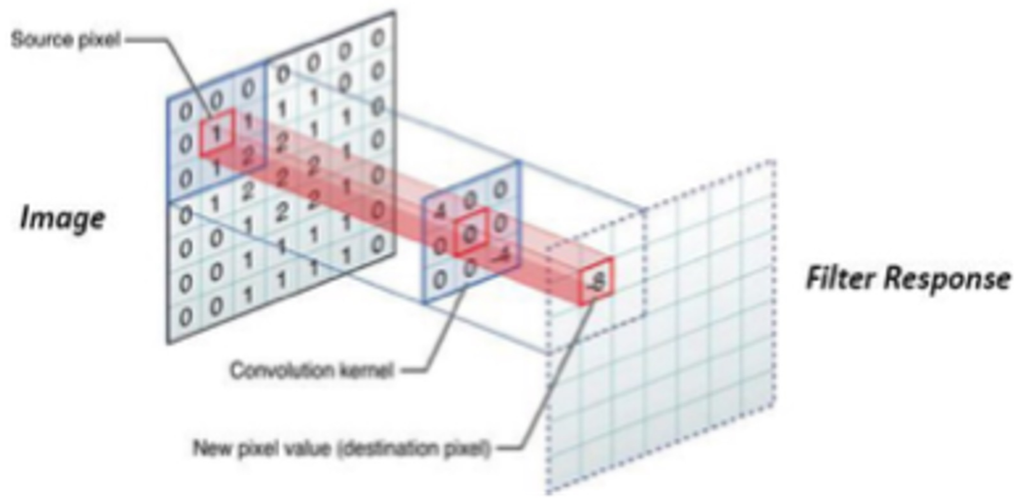
Convolutional Neural Networks

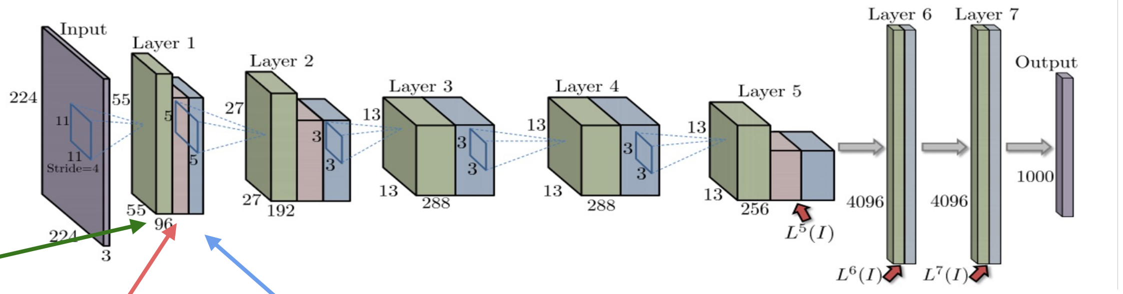




Activation functions

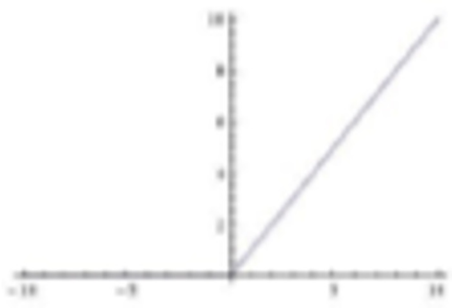
Convolutional Neural Networks



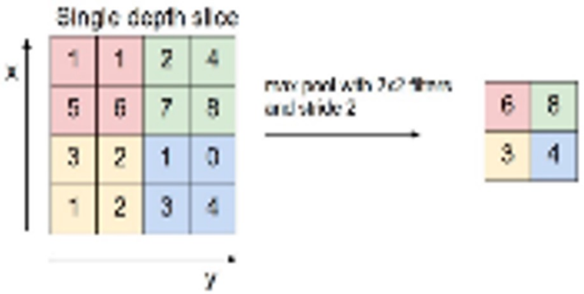


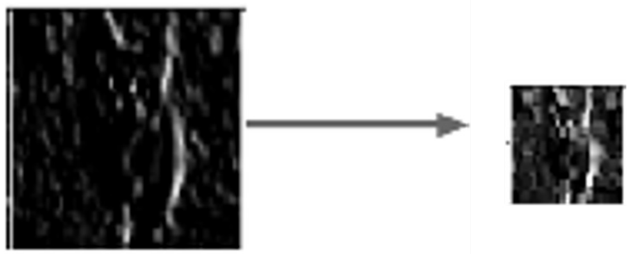
Convolutional layer

ReLU max(0,x)



Max pooling layer





Convolutional Neural Networks

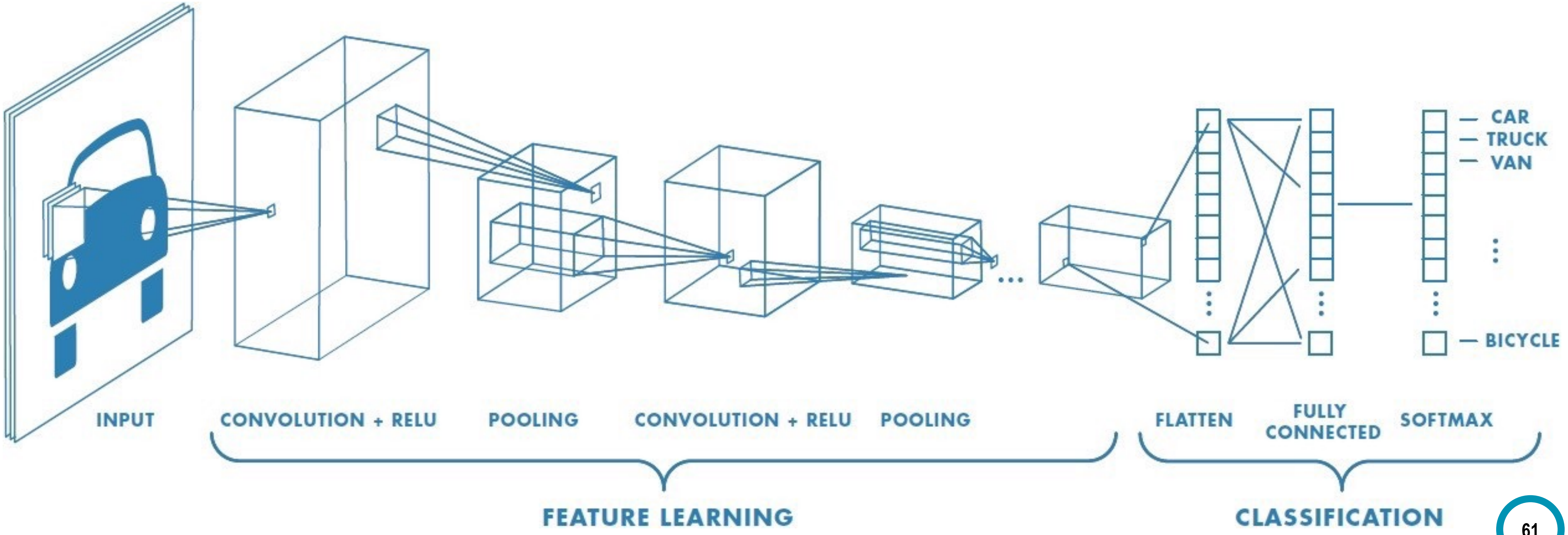
- Backpropagation and chain rule
- Loss functions (cross-entropy loss, binary loss, …)
- Optimization, gradient descent, momentum
- Etc

… Machine learning 1 and 2 and Deep learning courses! ☺

3Blue1Brown series on Deep Learning
https://youtu.be/aircAruvnKk

Convolutional Neural Networks

Possible outputs:
- Semantic classification
- Vectorial representation *can be used as image descriptors*



https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

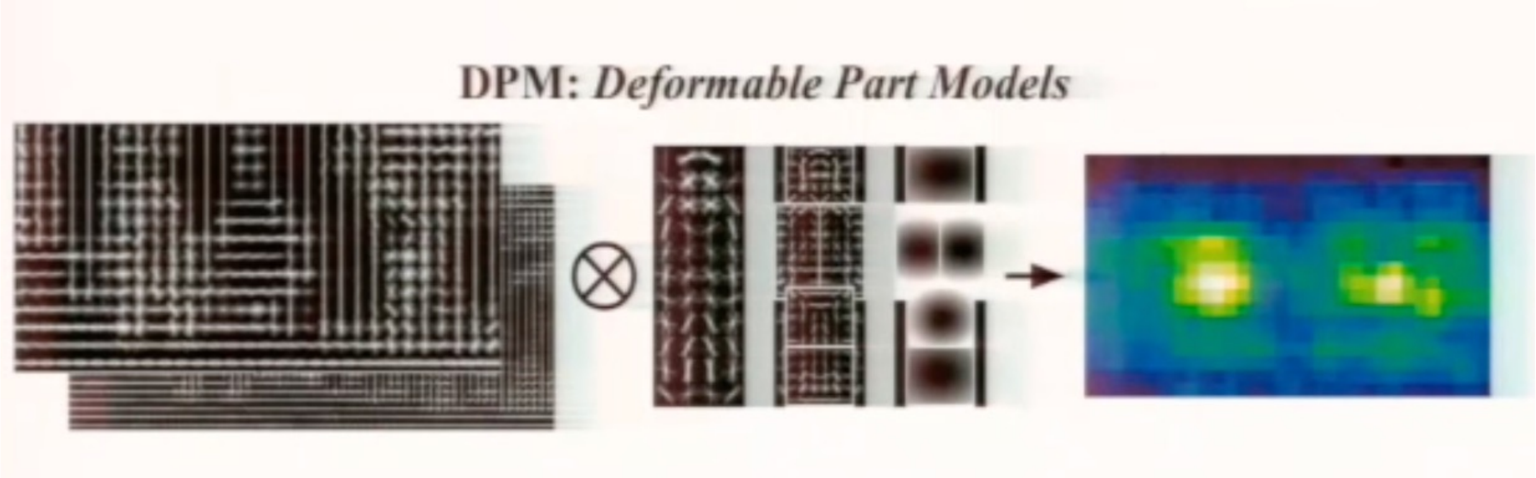
Use of detected semantics as descriptors:

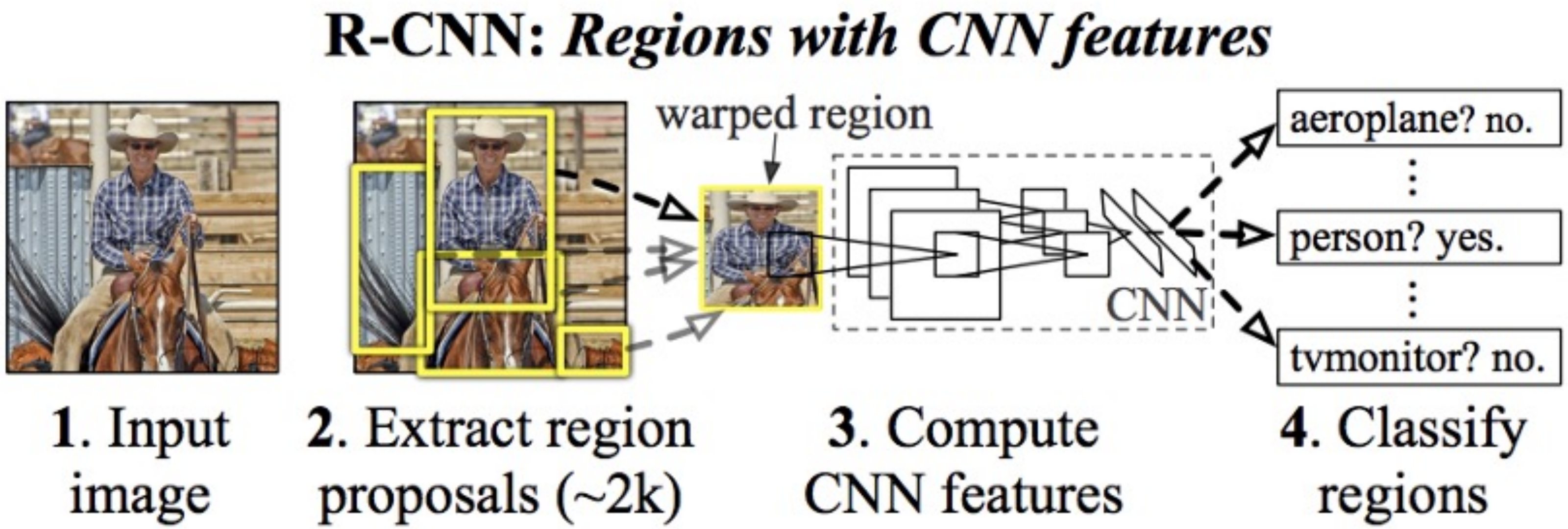
Object detection with CNNs

Region-based classifiers
Yolo

Reading: https://datascience.eu/computer-vision/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms/

Region-based Classifiers

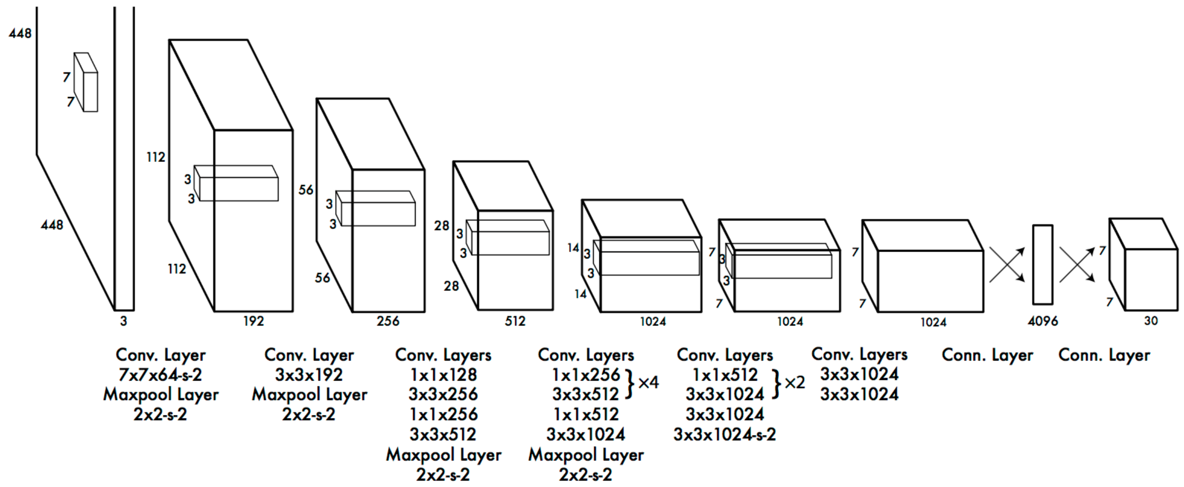


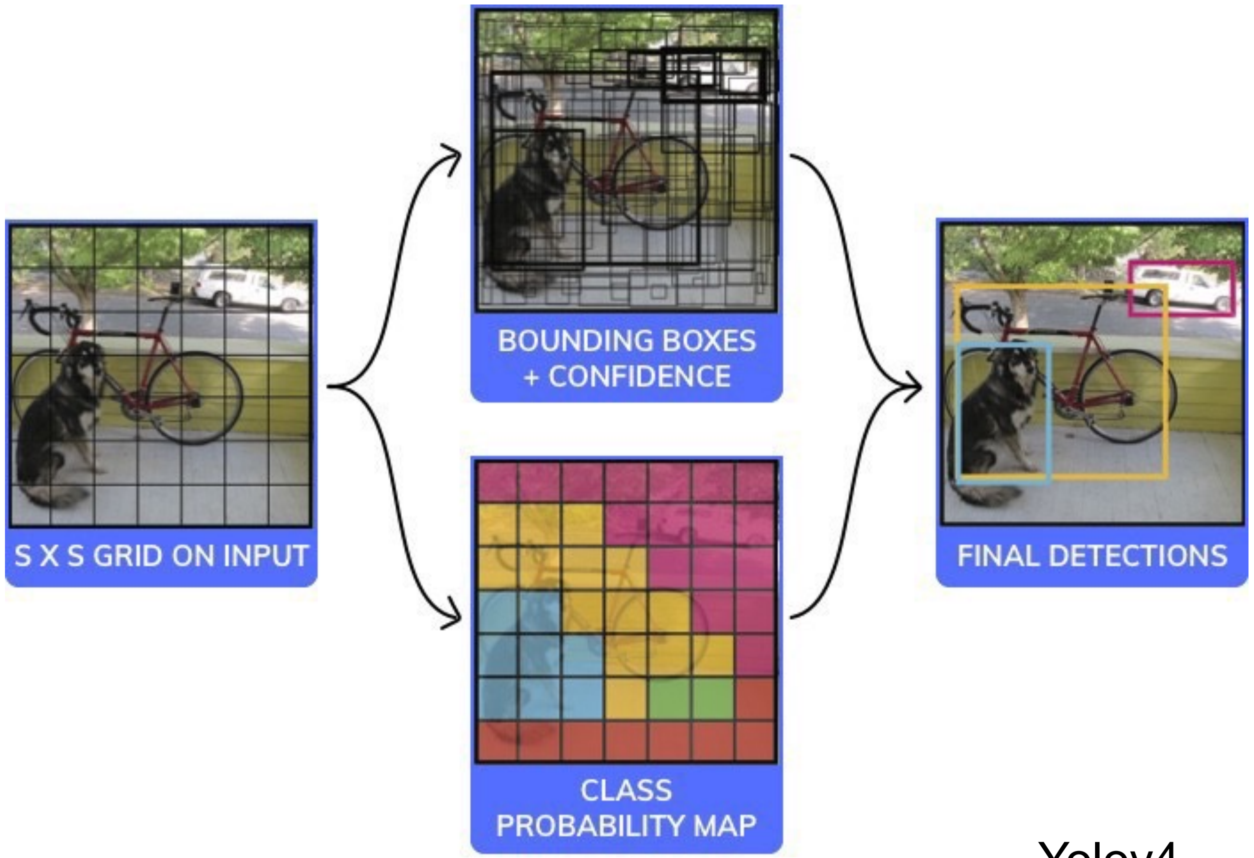


https://towardsdatascience.com/object-detection-explained-r-cnn-a6c813937a76

You Only Look Once: Unified, Real-Time Object Detection

https://www.youtube.com/watch?v=NM6Irxy0bxs





…Yolov4

Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

2 MACHINE LEARNING

GENERALIZATION & VALIDATION

Generalization is the ability of the model to react to unseen data samples.

The **validation** of a model is the process where a trained model is evaluated with a testing data set

Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.

key problems in supervised learning

model selection: Neural Networks, Support Vector Machine, Decision trees… ?
how many prototypes/seeds, neurons, which kernel,… ?

data representation: coding, normalization, transformation, … ?

algorithm, (hyper-) parameters:
which training prescription ?
how many training epochs, which learning rate, … ?

---

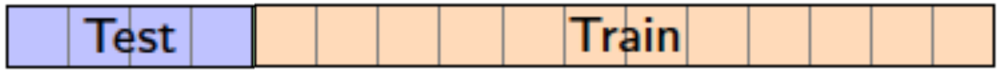
training: based on performance with respect to training data
aim : low error with respect to new data **generalization**

how can we test the generalization performance ?

Validation procedures

basic idea: split available data $D = \left\{ \left\{ \xi^\mu, S^\mu \right\} \right\}_{\mu=1}^{P}$
(randomly) into disjoint sets

$$D_{training} = \left\{ \left\{ \xi^\mu, S^\mu \right\} \right\}_{\mu=1}^{Q} \qquad D_{test} = \left\{ \left\{ \xi^\mu, S^\mu \right\} \right\}_{\mu=Q+1}^{P}$$



In practice, available training data is often limited

Splitting the data in sets further reduces this

→ estimate of test error E_{test} (e.g. number of misclassifications)

→ comparison/choice of different models, algorithms, parameter settings

→ prediction of performance with respect to novel data (?)

problems:

- **lack of data**

  can we afford to *waste* example data *only* for validation ?

- **representative results ?**

  lucky / unlucky set composition can give misleading outcome !

- **variation of results ?**

  how safe is the prediction ? error bars of the estimates ?

example strategy: " n-fold cross-validation "

split data $D = \left\{ \left\{ \xi^{\mu}, S^{\mu} \right\} \right\}_{\mu=1}^{P}$ (randomly) into **n** disjoint sets

$$D = \bigcup_{i=1}^{n} D^{(i)} \qquad D_{train}^{(i)} = D \,/\, D^{(i)} \qquad D_{test}^{(i)} = D^{(i)}$$

all data   training data (i)   test data (i)

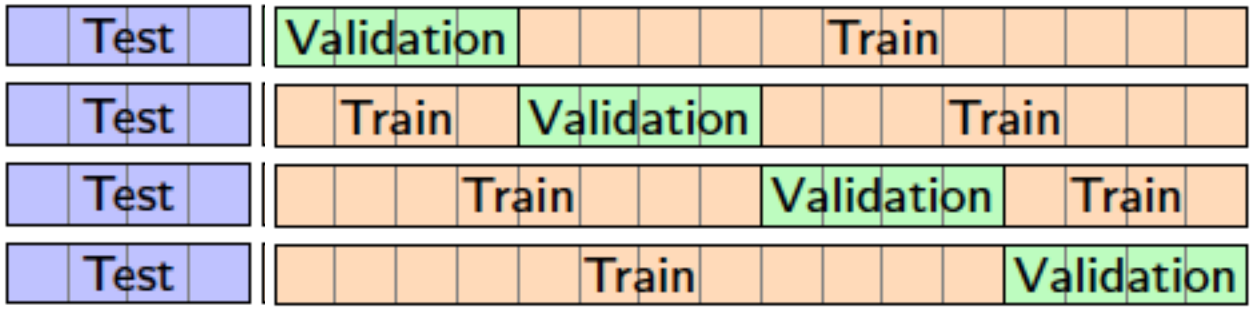
- repeat training n times
- calculate average training / test errors (and variances)

- repeat cross-validation for different models, parameter settings, etc.
- select the best system with respect to test errors

(model, number of units, learning rate, …)

n-fold cross-validation

Repeatedly split the data and average the results (here, n = 4)



- **which n** in n-fold cross-validation ?

larger n → larger fraction of D used in each training run

→ more estimates of E_{test} / smaller test sets

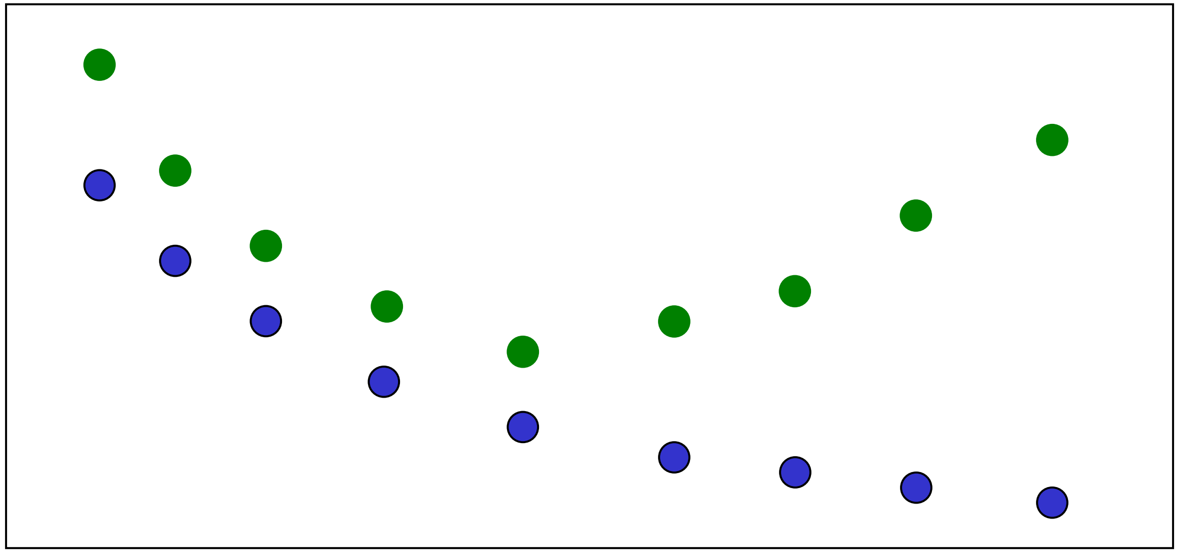
→ higher computational effort

extreme case: n = P,

repeat P times " **leave-one-out estimate** "

remarks:

test / training errors (e.g. observed in cross-validation)
vs. complexity of the model (e.g. # of prototypes, neurons, …)

E


in general:

$E_{train} < E_{test}$

"complexity“ (e.g.: number of prototypes, hidden units…)

- expect:
better classification (of D_{train}) with increasing complexity

Performance comparison *Validation metrics*

What error measure (or utility function)?

Different problems require different types of ***Error Functions***:

Supervised learning
Classification Some form of counting misclassified datapoints
Regression Average distance between predicted and target output

Unsupervised learning Within-class and between-class distances between samples

Performance in **Supervised Learning** - Classification

A confusion matrix		**Predicted classes**		
		Positive	Negative	Total
Groundtruth classes	Positive	TP	FN	POS_{GT}
	Negative	FP	TN	NEG_{GT}
	Total	POS_P	NEG_P	N

$$Acc = \frac{TP + TN}{N}$$

Additional class-specific metrics:

$TPR = \frac{TP}{POS_{GT}}$ (sensitivity or recall)

$$TNR = \frac{TN}{NEG_{GT}}$$

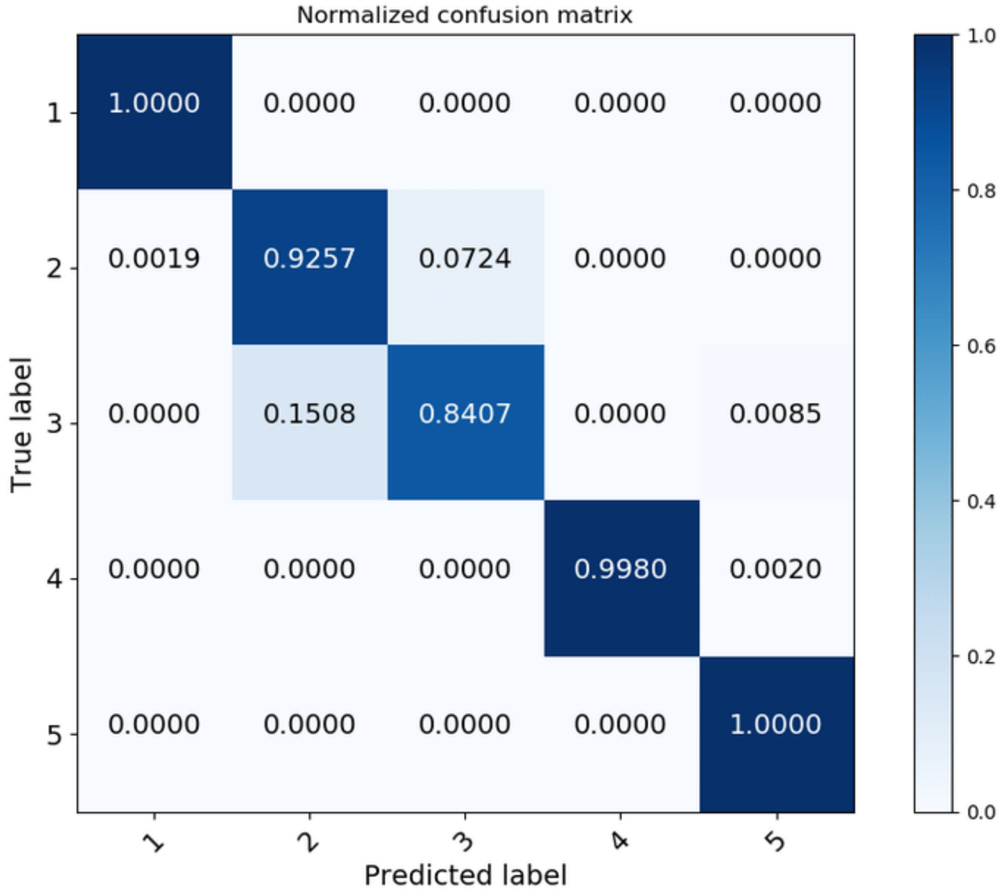
$$FNR = \frac{\dots}{\dots}$$

$$TNR = \frac{\dots}{\dots}$$

$$\text{Precision} = \frac{TP}{POS_P}$$

Performance in **Supervised Learning** - Classification

Multi class confusion matrix

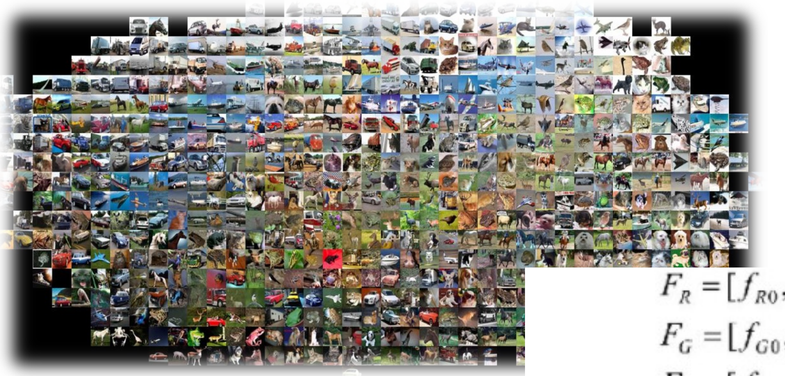


Recap

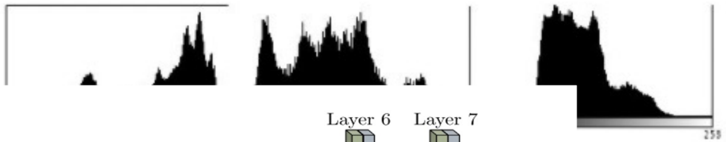
Brief intro to AI

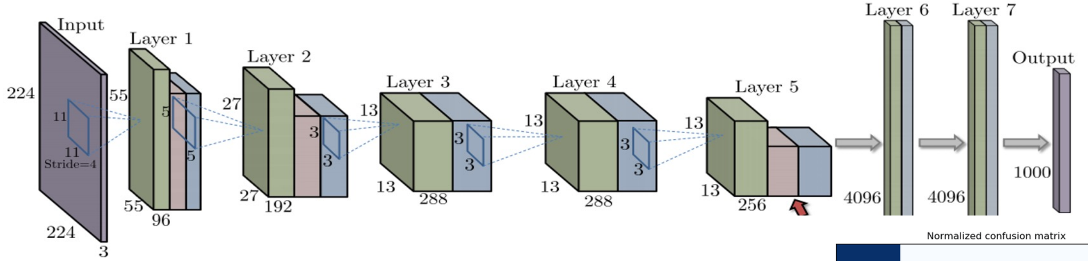
- Computer vision
- Image classification
  - Machine learning

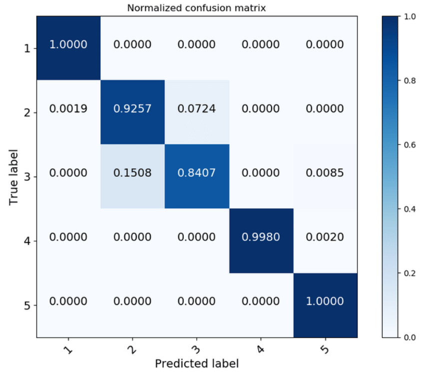
- Practicals
  - Assignments
  - Project



$$F_R = [f_{R0}, \dots, f_{R255}]$$
$$F_G = [f_{G0}, \dots, f_{G255}]$$
$$F_B = [f_{B0}, \dots, f_{B255}]$$
$$F_{RGB} = [F_R \mid F_G \mid F_B]$$







3 CIRI PROJECT

Assignments

main · 1 branch · 0 tags · Go to file · Add file · Code

estefaniatalavera Update README.md · c2bf31e 4 days ago · 19 commits

Assignment 0v1 - Skimage intro.ipy...	Add files via upload	4 days ago
Assignment 0v2 - Numpy intro.ipynb	Add files via upload	4 days ago
Assignment 0v3 - Image manipulati...	Add files via upload	4 days ago
Assignment 1 - Understanding a da...	Add files via upload	4 days ago
README.md	Update README.md	4 days ago
[Optional] Assignment 3 - Transfer ...	Add files via upload	4 days ago

README.md

Assignments DS-UT

This folder contains the material prepared in January 2022 for the master course Data Science at the University of Twente.

A) Set of Jupyter notebooks with examples on how to read images and work with them Numpy Intro, Skimage Intro, Basic Image manipulation.

B) Assignments 1 and 2 are meant to be solved sequentially and therefore are related. You will be selecting the same dataset(s) to solve both assignments.

CIRI Project Incident recognition in images

Goal:
Classificaton of incidents depicted in images

Dataset:
Subset of the Incident dataset [1]

Challenges:

- Design and implement ML framework for image classification
- Visualize and represent the data in an informative way
- Evaluate and discuss performance of the implemented models

This session gave you
tools to work on the project
from different perspectives,
Investigate them further and use them
to build something nice!

[1] Weber, Ethan, et al. "Detecting natural disasters, damage, and incidents in the wild." *European Conference on Computer Vision*. Springer, Cham, 2020.

CIRI Project Image classification framework

Input data



…



Images (+labels) + supervised ML = Image classification

Feature extraction → Feature processing → ML

Classification

- **0** Flooded
- **1** Airplane accident
- **2** Oil spill
- **3** Burned
- …
- **n-1** Thunderstorm
- **n** Volcanic Eruption

Images: Weber, Ethan, et al. "Detecting natural disasters, damage, and incidents in the wild." *European Conference on Computer Vision*. Springer, Cham, 2020.

CIRI Project Report

What to report to create a good report?

Dataset description (table describing the data split)

Performance of the evaluated models applying cross validation (k = 3 or 5)

Results as Avg $\pm$ std

Confusion matrix

Examples of correct and incorrect classifications

How to write a good CV&IC report? Tips and suggestions:

- The data set needs to be described and presented. *Described:* Through **text and adding a table** quantifying the categories. *Presented:* **Include sample images** that describe the categories in dataset. You should also **describe** the **dataset cleaning process** and the **data split implemented.**
- You are suggested to include a **figure of your proposed framework** (stages the data goes through) that will help you better preset it to the reader, in combination with the text description.
- To assess generalisation capabilities of your model, you are expected to run **k-fold cross validation.**
- Implement a set of metrics to assess performance, start with **accuracy, weighted accuracy, f-score, precision, and recall**.
- Presenting a **confusion matrix** will help you present and describe the misclassifications of your model.
- **Present examples** of correctly and incorrectly classified samples

DATA SCIENCE

COMPUTER VISION

Image classification

Estefanía Talavera Martínez
e.talaveramartinez@utwente.nl

Some Slides from Fei-Fei Li, Ju[illegible] Johnson, Serena Yeung
http://vision.s[illegible]rd.edu/teaching/cs231n/

COMPUTER VISION & DEEP LEARNING

‘How we’re teaching computers to understand pictures’

Video
[Click here]

Fei-Fei Li · Computer scientist
As Director of Stanford’s Artificial Intelligence Lab and Vision Lab

COMPUTER VISION & DEEP LEARNING

Progressive Growing of GANs for Improved Quality

Video
[Click here]

Karras, et al., *Progressive Growing of GANs for Improved Quality, Stability, and Variation*, ICLR 2018

COMPUTER VISION & DEEP LEARNING

New Ego4D dataset

[Video
[click here]](#)

Deep learning (impact)

The Deep Learning Revolution (by NVIDIA)

[Video](#)
[[Click here]](#)

Motion perception
Johansson’s experiments [‘70s]

Video
[Click here]