TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Module 1: Introduction

5LSM0: Convolutional neural networks for computer vision

Fons van der Sommen

Electrical Engineering / VCA research group

VCA Video Coding & Architectures research group

Module outline

- What is this course about?
- What can you do with it?
- A brief history of computer vision
- The deep learning revolution
- Computer vision taxonomy (likely incomplete and completely subjective)
- Course details & planning


Image license free
(www.pexels.com)


Image license free
(www.pexels.com)


Image license free
(www.pexels.com)

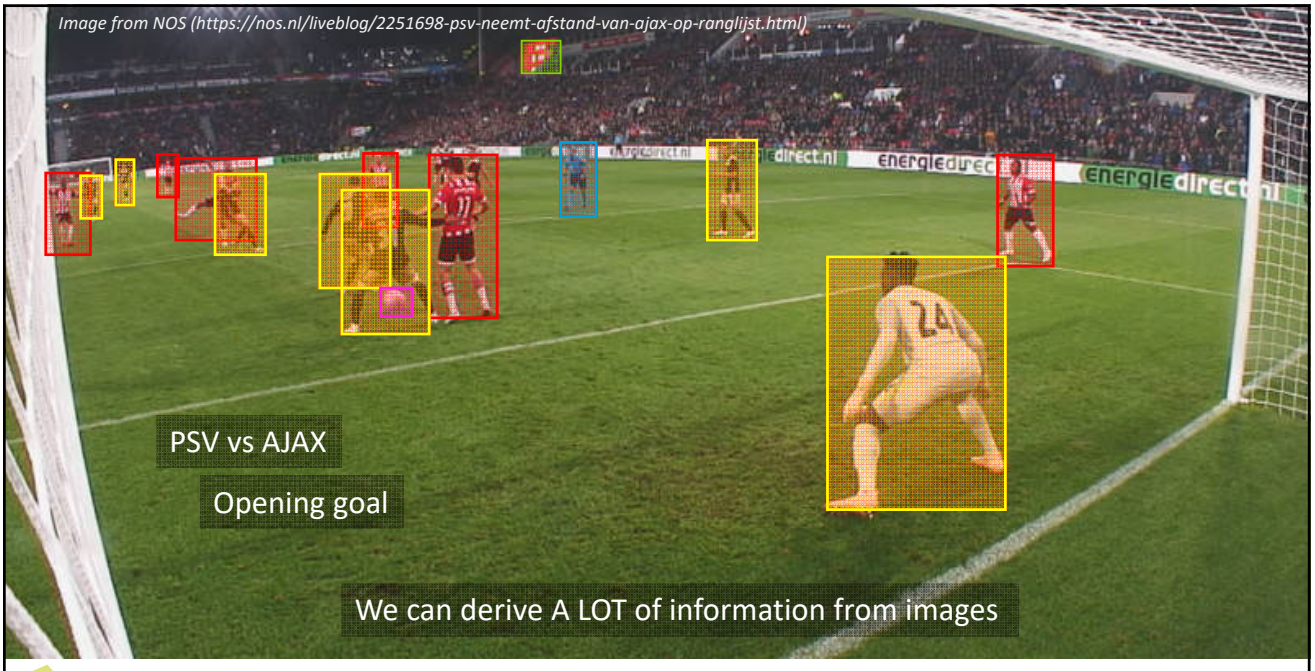


Image from NOS (https://nos.nl/liveblog/2251698-psv-neemt-afstand-van-ajax-op-ranglijst.html)
PSV vs AJAX
Opening goal
We can derive A LOT of information from images

What is this course about?

Image content analysis

Basically, a 3-year-old can do this…

Why is this hard for a computer?

- Enormous variation in appearance / illumination / viewpoint
  - *We cannot define a set of heuristic rules per object*
- Considerable variation in correct output
  - *Human language is very flexible and redundant*
- Work on a 2D mapping of a 3D world

What is this course about?

Computer vision

- Can we make a computer see? (Image understanding)
- Machine learning + image processing (+robotics/optics/algorithms/NLP/…)

Deep learning

- Revolutionized machine learning/computer vision
- Relatively old idea
- Use an hierarchical structure of filters
  - *Convolutional Neural Network (CNN)*

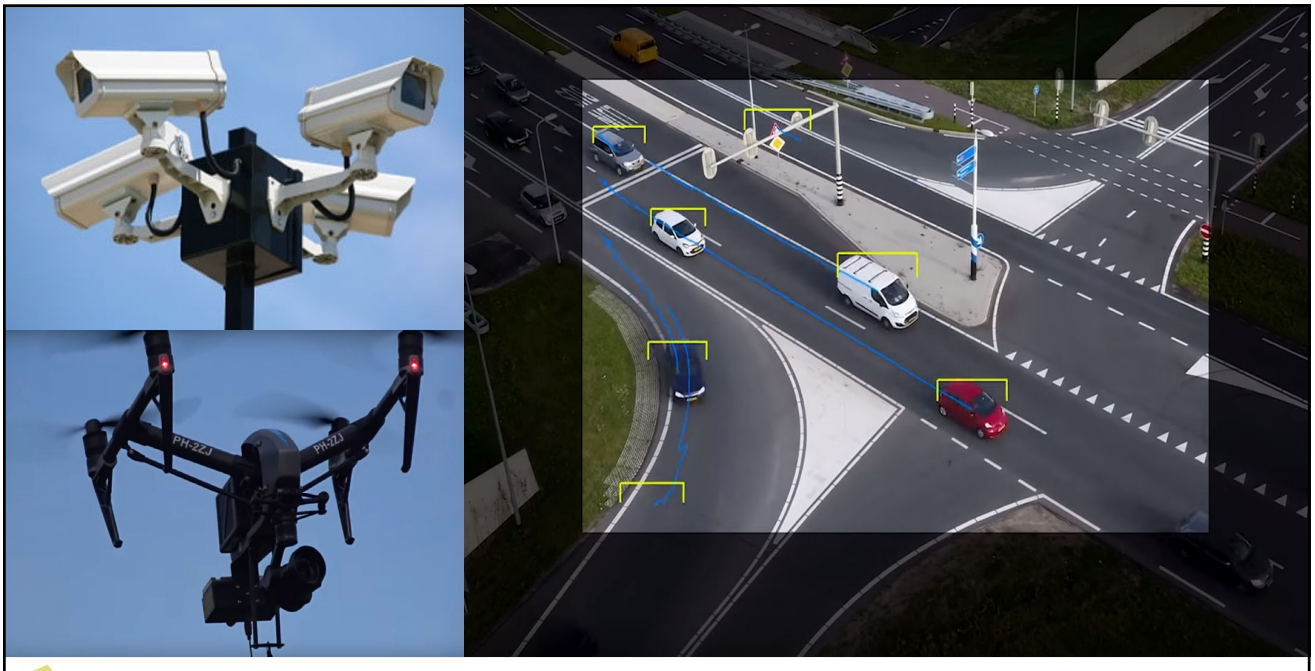
What can you do with it?

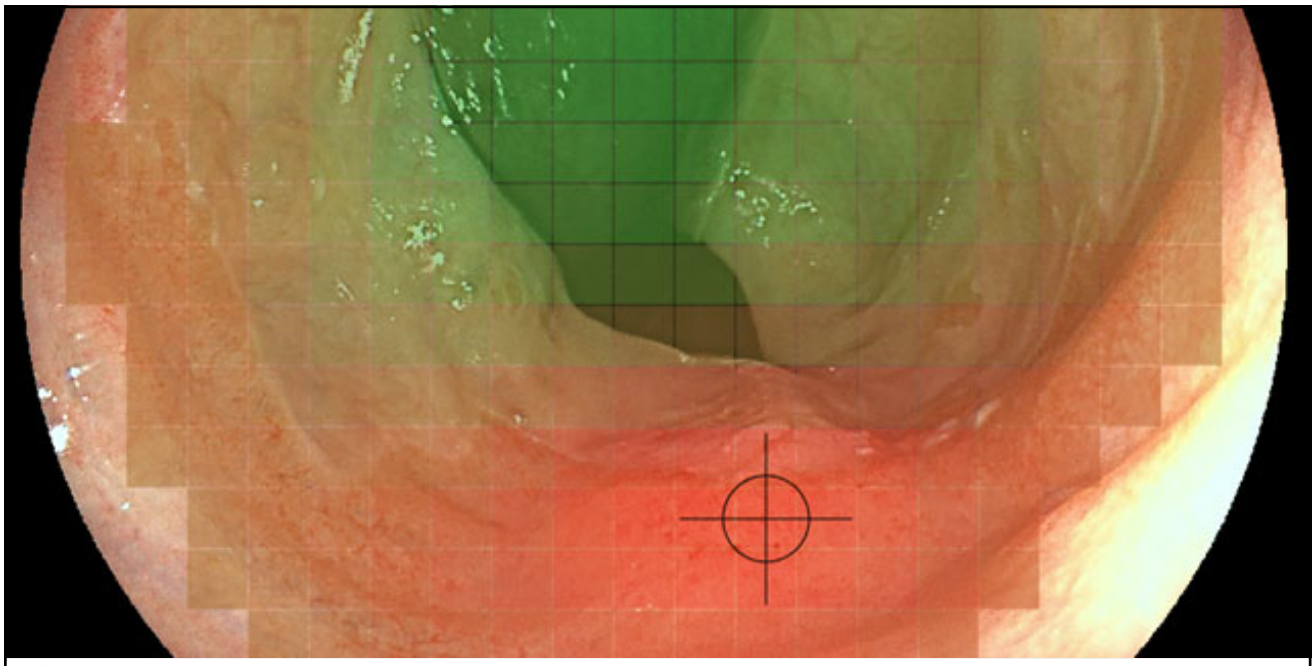
Amount of visual data is exploding

- 5 hours of video uploaded to YouTube every second
- 45 billion cameras in the world by 2022
- 81% of transmitted data on the web is visual

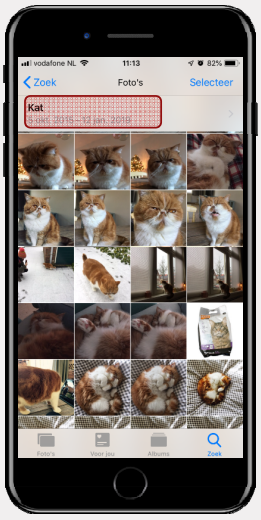
Growing number of application domains

- Surveillance
- Automotive
- Healthcare
- …



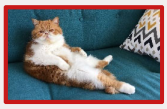


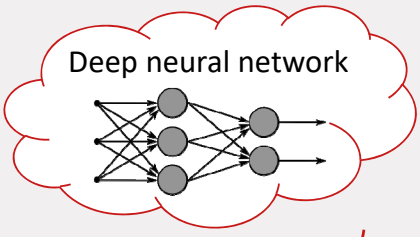
Search for "Cat"



But how???

I never labeled these images...





94%	**Cat**
3%	Couch
2%	Cushion
1%	…

Unthinkable <10 years ago…

Disclaimer

This course is loosely based on a course from Stanford university:

CS231n: Convoutional neural Networks for Visual Recognition

- http://cs231n.stanford.edu/
- Lectures of previous editions on YouTube!

However, we do not cover all topics addressed in CS231n and we also include topics in 5LSM0 that are not in CS231n.

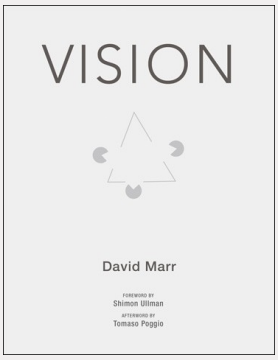
CS = Computer Science

We aim to approach the topic from a more Signal Processing point of view.

A brief history of computer vision

Early days

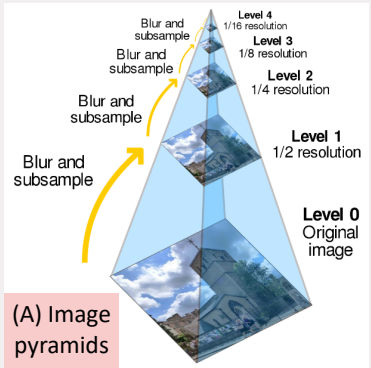
- Summer vision project @ MIT (1966), Seymour Papert
  - *10 students to do object detection*
- Line detection and line-labeling algorithms
- Elastic arrangement of "Pictorial structures" (Fischler & Elschlager, 1973)
- First feature-based stereo correspondence algorithms
  - *Use interest points to find matches in two images of the same object*
  - *Use camera parameters to compute 3D locations of those points*
- State-of-the-art nicely summarized in David Marr's "Vision" (1982)
  - *Different stages of vision (primal sketch, 2.5D sketch, 3D model)*

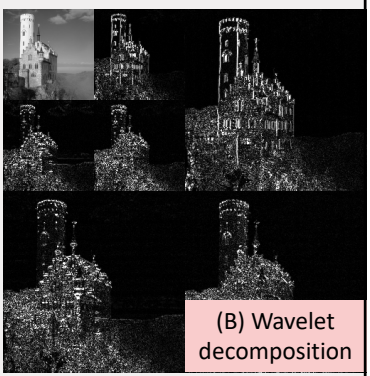


A brief history of computer vision

1980s

- Image pyramids
- Scale-space processing
- Wavelet decomposition
- Shape-from-X
- Active contours
- Variational optimization problems and regularization
  - *Markov Random Fields (MRFs)*





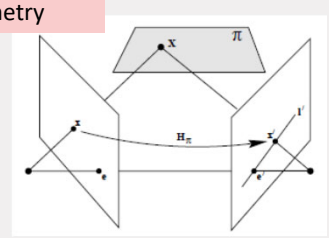
Images from Wikimedia Commons (A), (B), (C)

A brief history of computer vision

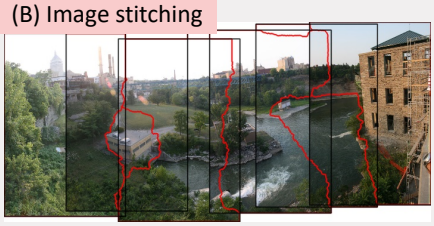
1990s

- Projective reconstructions using camera approximations
  - *Multi-view stereo algorithms*
- Color image processing
  - *Physical models of image acquisition*
- Optical flow
- Tracking
- Image segmentation
- Statistical learning
- Interaction with computer graphics -> image stitching

(A) Multiple view geometry



(B) Image stitching



(A) p243 in "Hartley, R., & Zisserman, A. (2003). Multiple view geometry in computer vision. Cambridge university press."

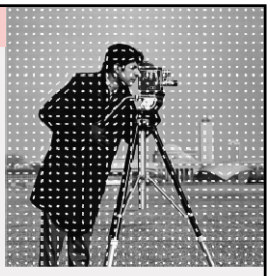
(B) Wikimedia Commons

A brief history of computer vision

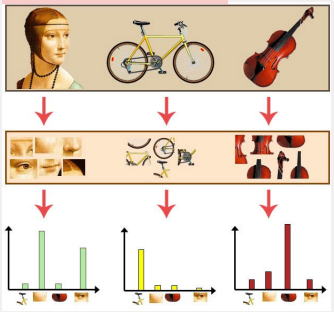
2000s

- Texture synthesis and inpainting
- Feature-based techniques combined with learning
- Scale-Invariant Feature Transform (SIFT)
- Speeded-Up Robust Features (SURF)
- Histogram of Oriented Gradients (HOG)
- Bag Of visual Words (BoVW)
- Face detection (Viola & Jones)

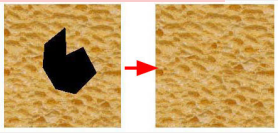
(B) HOG features



(C) BoVW



(A) Texture synthesis



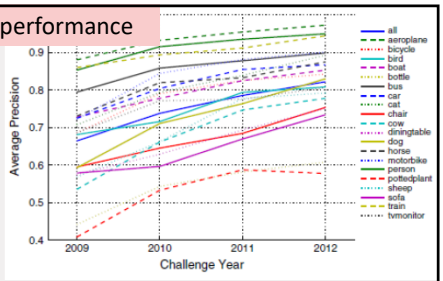
Images from MathWorks (A), (B), and towardsdatascience.com (C)

A brief history of computer vision

Increase of data and emergence of public challenges

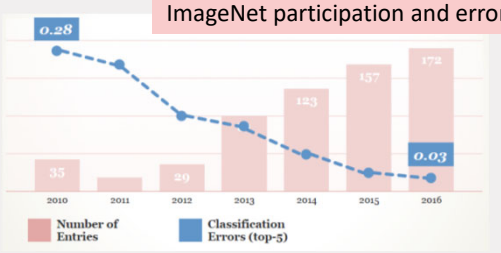
- PASCAL Visual Object Challenge[1]
  - *~10k images, 20 classes*
  - *Person, bird, cat, aeroplane, bicycle, boat, car, bottle, chair, potted plant, sofa, tv/monitor*
- ImageNet
  - *22K categories and 14M images*
  - *Large-Scale Visual Recognition Challenge (LSVRC)*
  - *1,000 classes and 1.4M images*

PASCAL performance



Everingham *et al.*, (2015). The pascal visual object classes challenge: A retrospective. IJCV, *111*(1), 98-136.

ImageNet participation and error



O. Russakovsky et al., Detecting avocados to zucchinis: what have we done, and where are we going?, *Proc. ICCV. 2013.*

1) Everingham et al. (2010). The pascal visual object classes (voc) challenge. *IJCV, 88*(2), 303-338.

The deep learning revolution

DEEP LEARNING IS TAKING OVER!!1111

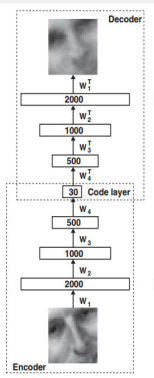
And it's coming for you!

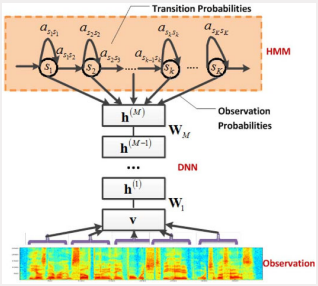
Or your data, at least…

A scene from "Terminator."
(Screenshot: Warner Bros. via CNET/CBS Interactive)

The deep learning revolution

Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "**Reducing the dimensionality of data with neural networks.**" *science* 313.5786 (2006): 504-507.





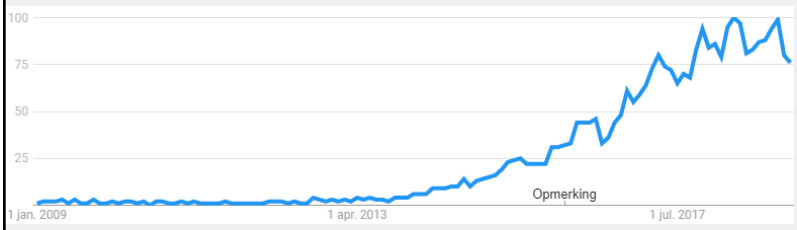
Context-Dependent Pre-trained Deep Neural Networks for Large Vocabulary Speech Recognition
George Dahl, Dong Yu, Li Deng, Alex Acero, 2010

Nature cover, January 2016

Science cover, July 2017

The deep learning revolution

Google trends



"Convolutional Neural Network"

"Deep Learning"

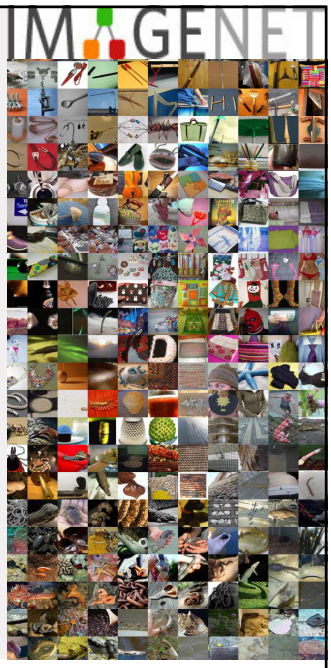
The deep learning revolution

Massive labeled datasets: ImageNet

- ImageNet Large Scale Visual Recognition Challenge (ILSVRC)
- 1.2M images, 1000 categories

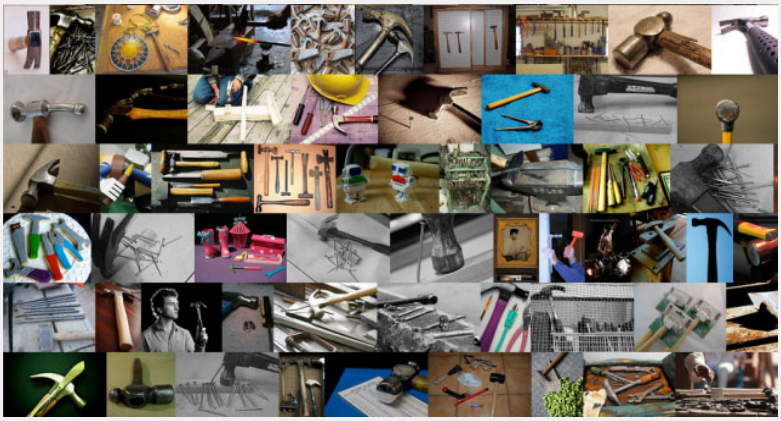
Inceased hardware capabilities

- Very powerful GPU's



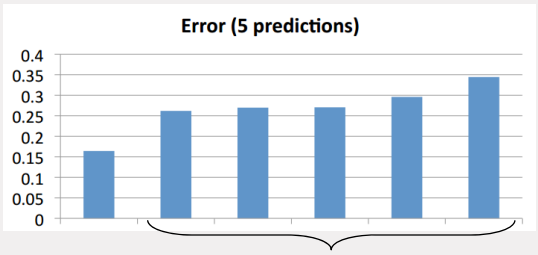
The deep learning revolution

Some test images for "Hammer" class



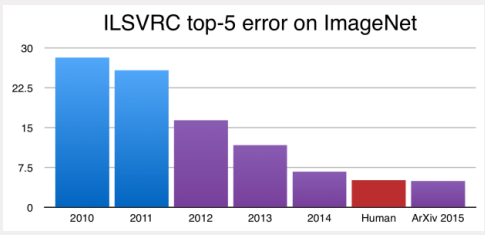
The deep learning revolution

First breaktrough in computer vision: 2012 ILSVRC won by a ConvNet



Top5 error =16.422%

Top5 error >26.172%



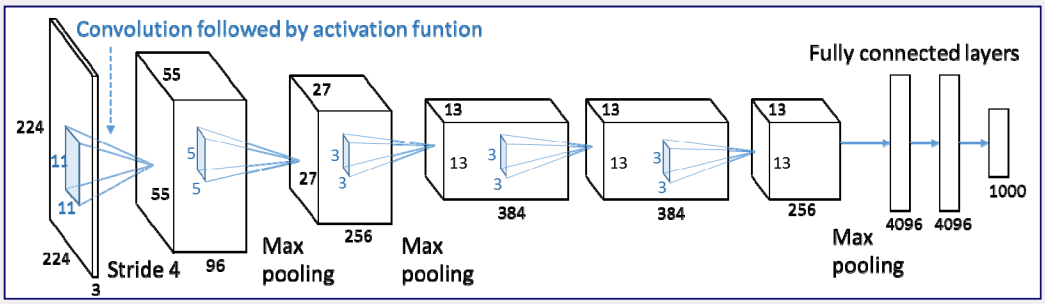
ImageNet error over time

The deep learning revolution

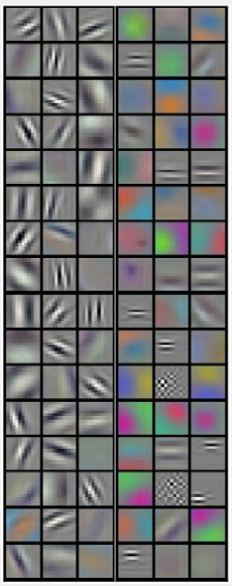
2012 ILSVRC winner: AlexNet

- Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton: ImageNet Classification with Deep Convolutional Neural Networks. NIPS 2012: 1106-1114.

AlexNet architecture

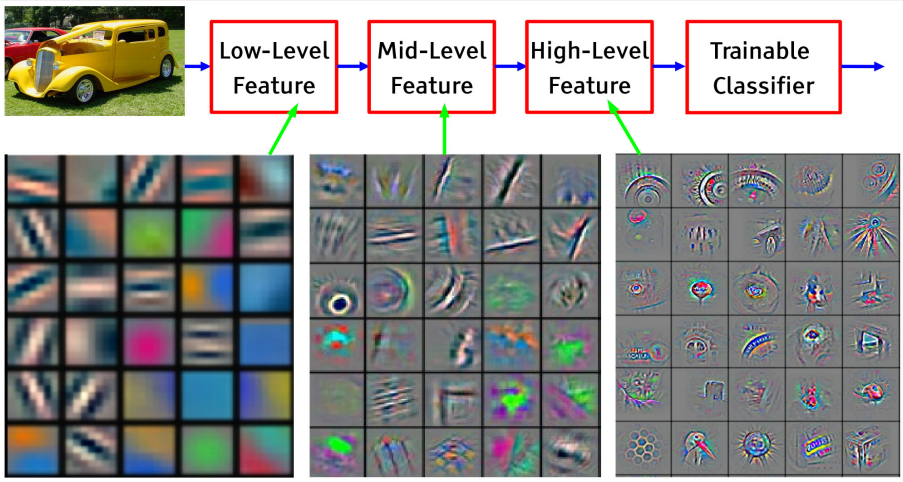


Convolution kernels of the first layer



The deep learning revolution

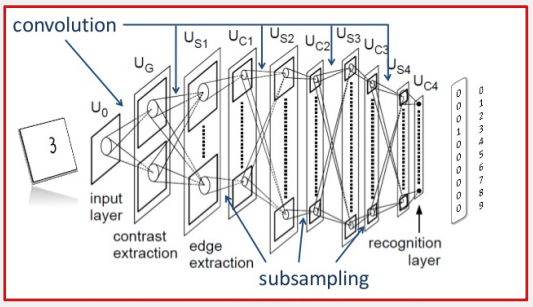
Hierarchical representation



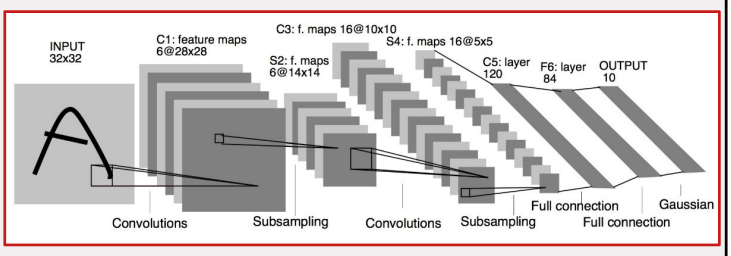
Courtesy Zieler and Fergus, ECCV, 2013

The deep learning revolution

The idea is actually pretty old

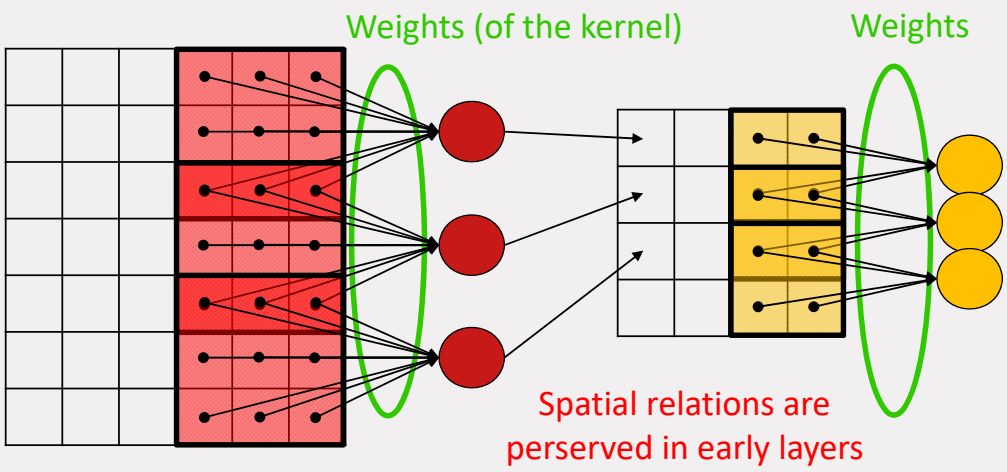


Fukishima, 1980, Neurocognition



Lecun, 1998, LeNet-5

The deep learning revolution



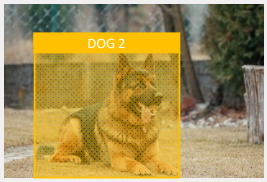
The deep learning revolution

- How can we find these filter kernels and the optimal hierarchical structure?
- What are the key ingredients of a ConvNet architecture?
- What do we need to train such a network?
- How can we evaluate the performance of such such networks?
- What image-understanding problems can we solve using convnets and how?
- What are the important parameters that affect their succes and how do we pick them?
- Can we also apply ConvNets if we don't have sufficient data to train a deep network?
- Can we also apply deep learning if we don't have a ground truth?
- What do ConvNets learn?

… all questions that you will be able to answer after this course!

Images license free (www.pexels.com)

Computer vision taxonomy



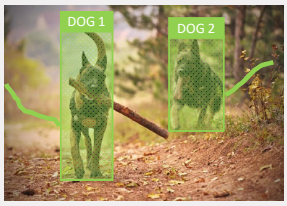
Re-identification

Classification

Segmentation

Tracking

Detection



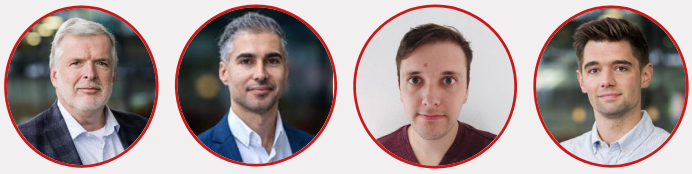
Course details & planning

Lectures + computer classes

- Effectively 8 weeks
- Lectures on Tuesday 5+6 [Helix West 1.91] and Thursday 1+2 [Atlas 4.225]
- Computer class on Tuesday 7+8 [Helix Oost 4.91] and Thursday 3+4 [Atlas 4.225]
- First weeks: Python tutorial during computer class
  - *Strongly advised*

People *(f.l.t.r)*

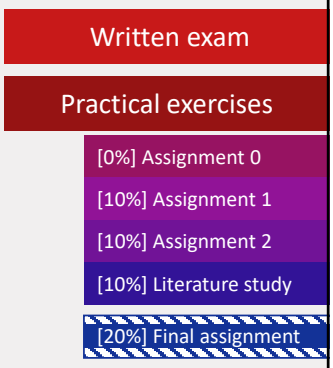
- Peter de With
- Farhad Ghazvinian Zanjani
- Joost van der Putten
- Fons van der Sommen



Course details & planning

Exam & grading

- Default: Written exam (70%) and assignments (30%), if possible →
- Allowed to work in couples for assignments, but individual reports



Materials

- Slides are sufficient to study for the exam
- Optional book on deep learning:
  - *Goodfellow, Bengio & Courville, "Deep Learning" (2015)*
  - *Free online! (https://www.deeplearningbook.org/)*
- Book on conventional computer vision (background)
  - *Richard Szeliski, "Computer Vision, Algorithms and Applications" (2011)*