

Neural Networks for Computer Vision

Lecture 1: Introduction

Ing. Viktor Kocur, PhD., RNDr. Zuzana Černeková, PhD.

22.9.2021

Contents





- Computer Vision
- History of CV
- Deep Neural Nets
- CV Applications
- Al Hype
- Syllabus
- Grading
- Recommended Literature

Computer Vision





Computer Vision

CV is an interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images or videos.

Computer Vision



Computer Vision

CV is an interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images or videos.

Understanding

Understanding in this context means the transformation of visual images into descriptions of the world that make sense to thought processes and can elicit appropriate action.

Two approaches



Engineering

From the perspective of engineering, it seeks to understand and automate tasks that the human visual system can do.

Two approaches



Engineering

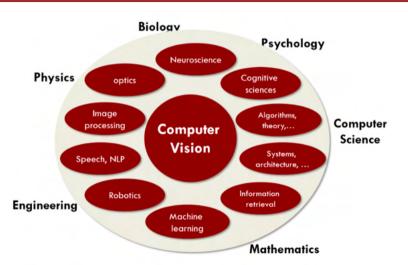
From the perspective of engineering, it seeks to understand and automate tasks that the human visual system can do.

Science

The scientific discipline of computer vision is concerned with the theory behind artificial systems that extract information from images.

CV is interdisciplinary





Computer Vision and Computer Graphics



CV: Visual Information \rightarrow Abstract Representation

 $\textbf{CG: Abstract Representation} \rightarrow \textbf{Visual Output}$

Visual data



- There is an abundance of visual data
- Variety of sensors
- 80 % of internet traffic is video
- YT 5 hours of video get uploaded every second

Fundamental CV Task - Object Recognition





Fundamental CV Task - Object Recognition





 $\to \mathsf{dog}$

Biological Visual Systems

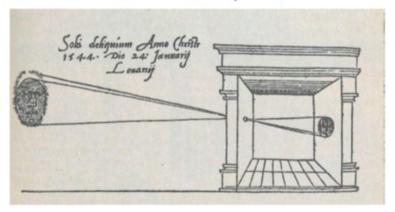


- Over millions of years vision has evolved to be the greatest sensory system in most animals.
- 50 % of human neurons are related to visual processing.
- Our visual system helps us survive, work, move, use tools, communicate, entertain ourselves and much more...





Gemma Frisius, 1545



Camera obscura - Ibn al-Haytham (11th century)

History - Image Capture





Joseph Nicéphore Niépce (1825)

History - Digital Imaging

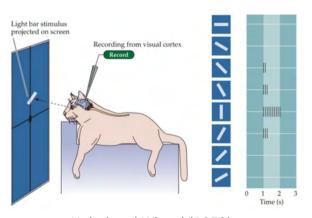




- Metal-oxide-semiconductor (MOS)
 - Mohamed M. Atalla and Dawon Kahng at Bell Labs
 - **1959**
- Charged-coupled device (CCD)
 - Willard S. Boyle and George E. Smith at Bell Labs
 - **1969**
 - Nobel prize in 2009
- Complementary MOS (CMOS)
 - Eric Fossum's team at the NASA Jet Propulsion Lab
 - **1993**

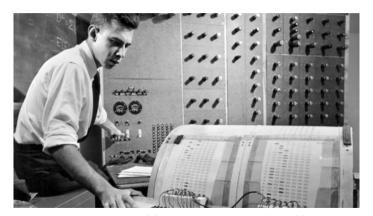
Early experiments





Hubel and Wiesel (1959)





Frank Rosenblatt - Perceptron (1960)

Computer Vision a Summer Project - 1966



MASSACHUSETTS INSTITUTE OF TECHNOLOGY PROJECT MAC

Artificial Intelligence Group Vision Memo. No. 100.

July 7, 1966

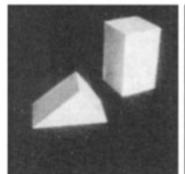
THE SUMMER VISION PROJECT

Seymour Papert

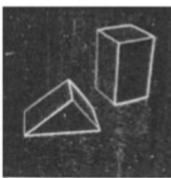
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

First CV PhD Thesis

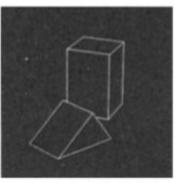




Input image



2x2 gradient operator

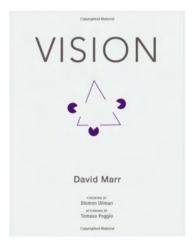


computed 3D model rendered from new viewpoint

Lawrence Roberts (1963)

Human Vision from Computational Perspective

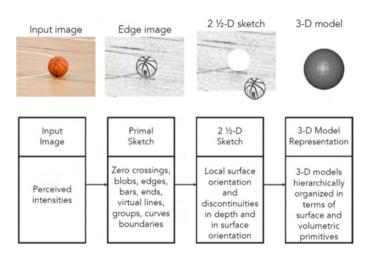




David Marr (1970s)

Human Vision from Computational Perspective





Local Features SIFT











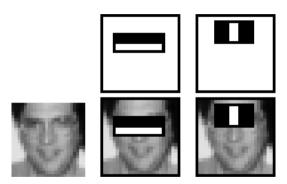


Davig G. Lowe (1999)

Image adopted from: David G Lowe. "Object recognition from local scale-invariant features." In: *Proceedings of the seventh IEEE international conference on computer vision*. Vol. 2. leee. 1999, pp. 1150–1157

Face Detection - Machine Learning





Viola and Jones (2001)



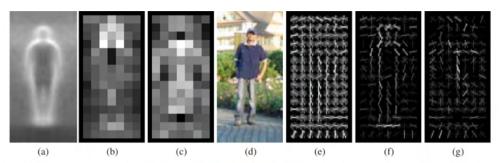


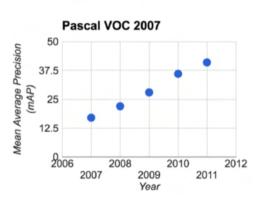
Figure 6. Our HOG detectors cue mainly on silhouette contours (especially the head, shoulders and feet). The most active blocks are centred on the image background just *outside* the contour. (a) The average gradient image over the training examples. (b) Each "pixel" shows the maximum positive SVM weight in the block centred on the pixel. (c) Likewise for the negative SVM weights. (d) A test image. (e) It's computed R-HOG descriptor. (f,g) The R-HOG descriptor weighted by respectively the positive and the negative SVM weights.

Dalal and Triggs (2005)

Pascal Visual Object Challenge



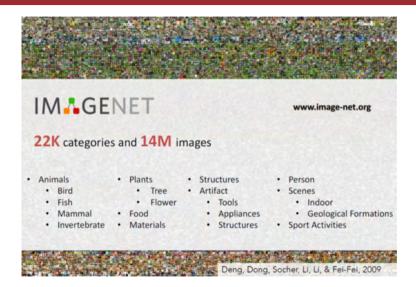




Everingham et al. 2006-2012

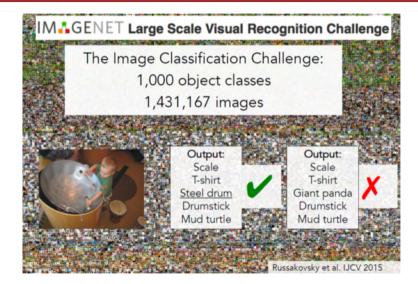
ImageNet





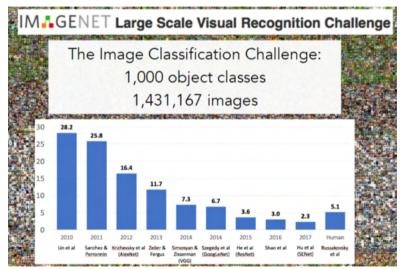






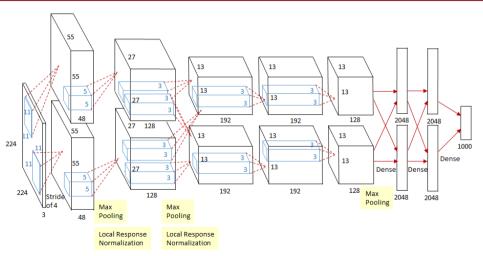






AlexNet

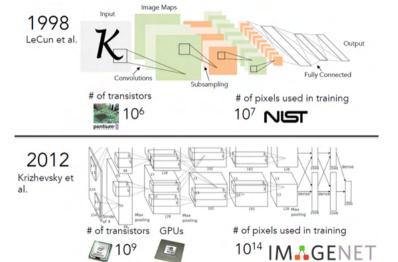




Krizhevsky, Sutskever and Hinton (2012)

AlexNet





Convolutional Neural Networks





- Replace human-engineered features (SIFT, HOG, etc.) with learned low level features
- Possible to use on different task with only the training data replaced
- Higher-level features emerge in later layers of the network due to training
- Convolution enables parameter-sharing and results in translational equivariance

Why CNNs became viable





Necessary conditions:

- More quality images available → large datasets
- GPU computing power

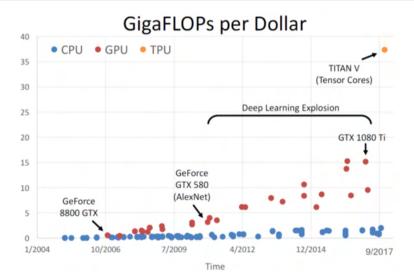
To win Imagenet AlexNet also employed:

- Custom CUDA implementation
- Splitting computation to two branches on separate GPUs
- Augmentation and preprocessing tricks

Evolution of Computational Efficiency







CV Tasks - Object Detection



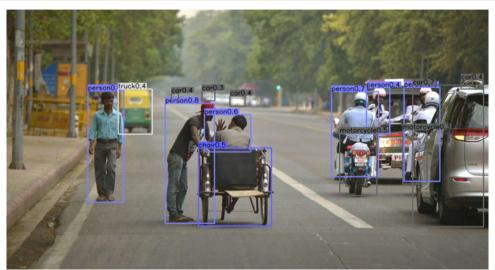


Image adopted from: Xingyi Zhou. CenterNet Github repo. https://github.com/xingyizhou/CenterTrack

CV Tasks - Object Detection





Image adopted from: Viktor Kocur and Milan Ftáčnik. "Detection of 3D bounding boxes of vehicles using perspective transformation for accurate speed measurement." In: Machine Vision and Applications 31.7 (4), pp. 1–15

CV Tasks - Object Detection



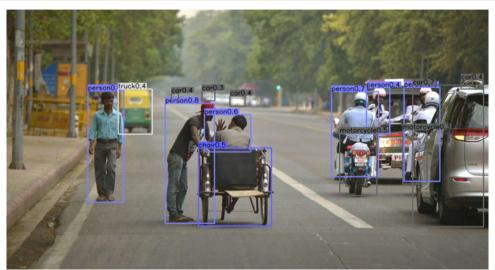
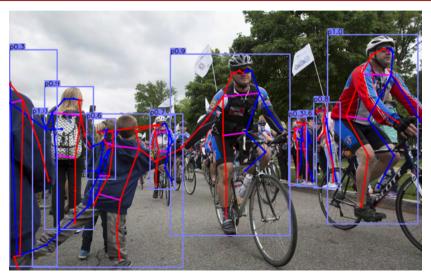


Image adopted from: Xingyi Zhou. CenterNet Github repo. https://github.com/xingyizhou/CenterTrack

CV Tasks - Pose Estimation





CV Tasks - Pose Estimation



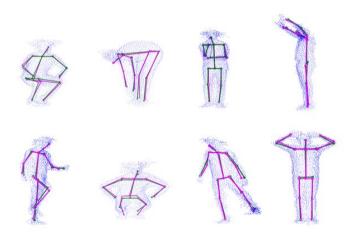


Image adopted from: Dana Skorvankova and Martin Madaras. "Human Pose Estimation using Per-Point Body Region Assignment." In: Computing and Informatics 32 (July 2021), pp. 1001–1020

CV Tasks - Instance Segmentation



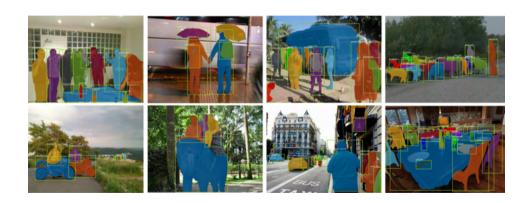


Image adopted from: Kaiming He et al. "Mask r-cnn." In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2961–2969

CV Tasks - Depth Estimation



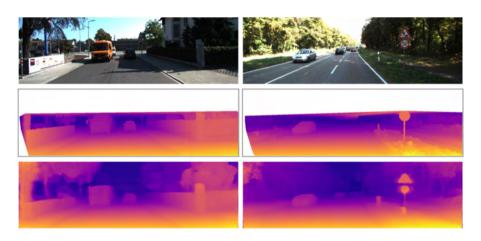


Image adopted from: Clément Godard, Oisin Mac Aodha, and Gabriel J Brostow. "Unsupervised monocular depth estimation with left-right consistency." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 270–279

CV Tasks - Sentence Generation







{aeroplane,fly,airport,at} the aeroplane is flying at the airport.



{person,motorbike,ride,field,in} the person is riding the motorbike in the field.



{person,bicycle,ride,street,on} the person is riding the bicycle on the street.



{person.table.sit.room.in} three people are sitting at the table in the room.

Style Transfer





Image adopted from: Leon A Gatys, Alexander S Ecker, and Matthias Bethge. "A neural algorithm of artistic style." In: arXiv preprint arXiv:1508.06576 (2015)

Image Generation



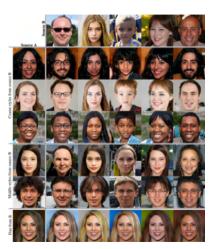


Image adopted from: Tero Karras, Samuli Laine, and Timo Aila. "A style-based generator architecture for generative adversarial networks." In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019, pp. 4401–4410

Image Generation



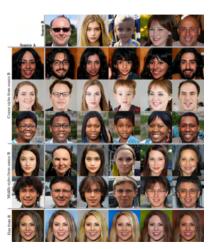


Image adopted from: Tero Karras, Samuli Laine, and Timo Aila. "A style-based generator architecture for generative adversarial networks." In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019, pp. 4401–4410

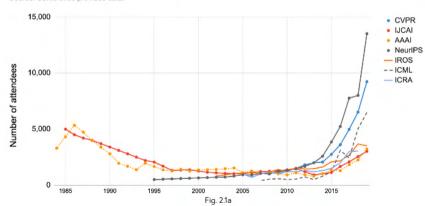
Deep Learning Revolution





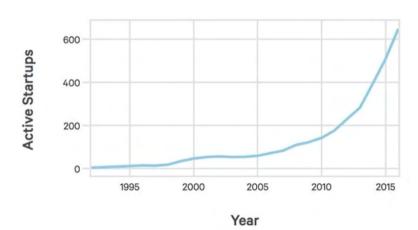
Attendance at large conferences (1984-2019)

Source: Conference provided data.



Al Startups

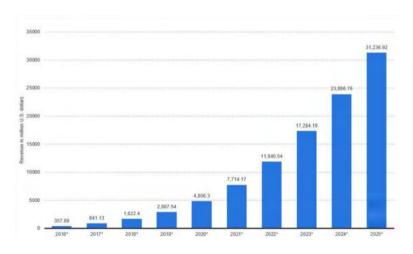




Al Enterprise Revenue







Are we close to true image understanding?





Image adopted from: Andrej Karpathy. The state of Computer Vision and Al: we are really, really far away. https://karpathy.github.io/2012/10/22/state-of-computer-vision/

What about classical CV





Classical CV methods are still usable!

- Insufficient/expensive data
- 3D Vision
- Image preprocessing
- Vision in enterprise solutions
- Domain specific problems
- Combination with DL

Grading



The course will be graded with an oral examination!

Syllabus I





- Introduction
- Learning, training and optimization
 - training, testing and validation sets
 - gradient optimization, loss functions
- Intro to neural nets
 - fully-connected neural nets
 - backpropagation
 - SGD
- Convolutional neural nets
 - convolution, pooling
 - activation functions

Syllabus II



- CNN Training I
 - vanishing/exploding gradients
 - batchnorm, dropout, normalization
 - hyperparameter tuning
- CNN Training II
 - Ada, Adam, Adagrad
 - augmentation
 - transfer learning
- Architectures
 - ► AlexNet, VGG, GoogLeNet
 - ResNets, DenseNets
 - ► MobileNets, EfficientNets
 - visual transformers

Syllabus III



- Recurrent neural nets
 - LSTM, GRU
 - transformers
 - captioning, query answering
- Object detection and instance segmentation
 - RetinaNet, YOLO, CenterNet, ATSS
 - ► Faster RCNN, Mask RCNN
- Generative models
 - ► VAE, GAN
 - training data generation

Syllabus IV



- Visualizing DL nets
 - filters, activation patches, saliency maps
 - DeepDream, adversarial attacks
 - textures, style transfer
- Ethics and science of deep learning
 - algorithmic bias
 - mass data collection
 - model robustness, ablation experiments

Recommended literature





- Ian Goodfellow et al. Deep Learning, MIT Press http://deeplearningbook.com
- Michael Nielsen Neural Networks and Deep Learning http://neuralnetworksanddeeplearning.com
- CS231n Stanford Course http://cs231n.stanford.edu
- Adrian Rosebrock Computer vision and deep learning, Resource guide
- Charu C. Aggarwal Neural Networks and Deep Learning: A Textbook

Related courses





- 2-AIN-269/19 Deep learning for computer vision labs
- 2-AIN-147/19 Computer Vision
- 2-AIN-233/00 Computer Vision Applications
- 2-AIN-112/15 Advanced Image Processing
- 2-AIN-204/10 Pattern Recognition
- 2-AIN-223/15 Virtual and Extended Reality
- 2-AIN-132/15 Neural Networks
- 2-INF-150/15 Machine Learning
- 2-INF-188/17 Current Approaches in Machine Learning