

Multimedia Systems

Luka Čehovin Zajc

About the lecturer



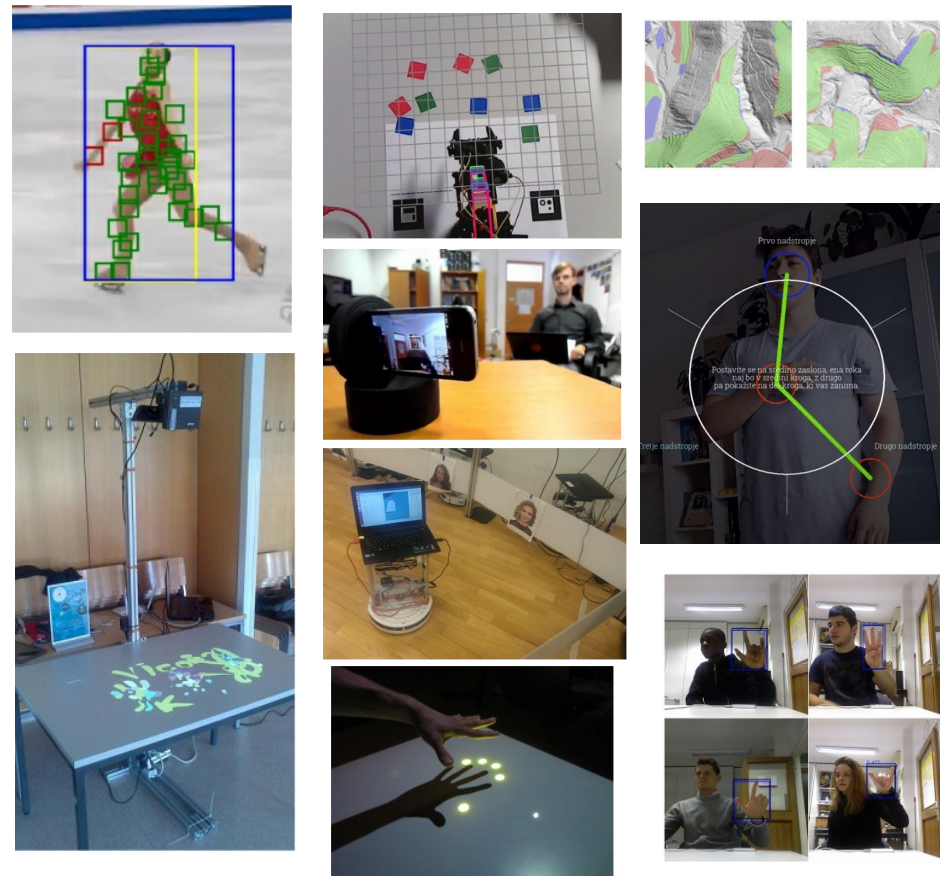
Luka Čehovin Zajc, PhD

Assistant Professor

Visual Cognitive Systems Laboratory

Room R2.35 (office hours online)

luka.cehovin@fri.uni-lj.si



Course requirements

- **Laboratory exercises / project work - 50%**
 - Practical exercises - grading throughout semester
 - Single project – grading at the end of the semester
 - Only valid for the current school year
- **Exam (written + oral) - 45%**
 - Must pass laboratory exercises to attend
 - Theoretical and practical assignments
 - Optional oral exam for borderline students (50% to ~65%)
 - Only oral exam for less than ~10 students
- **Lecture attendance – 5%**
 - Have to attend at least five lectures (not counting first one)

Laboratory exercises

- Teaching assistant: **Me**
- Practical consolidation of selected topics
- Python (Jupyter, SciKit, NumPy, ...)
 - Hosted Jupyter instances at lab.vicos.si
 - Local installation (virtualenv, Docker)
 - Google Colab
- Each exercise is due in **two weeks** (approximately)
 - **Timely** assignment hand-in encouraged
 - Stick to your designated **laboratory cycles** (for defenses!)

Project assignment

- Alternative to regular laboratory exercises
- Indepth project work on a selected topic
 - You have to pace your work yourself
 - Meetings can be arranged to discuss topic
- Work has to be finished by the end of semester
 - Presentation in classroom
 - Demonstration
 - Code hand-in
- Possible projects
 - 3D video stabilization using SfM
 - Content-based image retrieval with sketches
 - Content-based music retrieval in practice
 - Augmented reality without markers
 - Interactive / multitouch surfaces
 - Embedded devices for natural interaction



Write me an email
 if you are interested!

multimedia (*Latin*)
multum + medium

Different meanings

- **Computer salesman**
PC with GPU, sound-card, Blu-ray player, speakers?
- **Entertainment industry**
interactive digital TV with Internet connection, video on-demand
- **Computer science researchers / students**
interactive applications that utilize multiple modalities, text, images, animation, sound, etc.

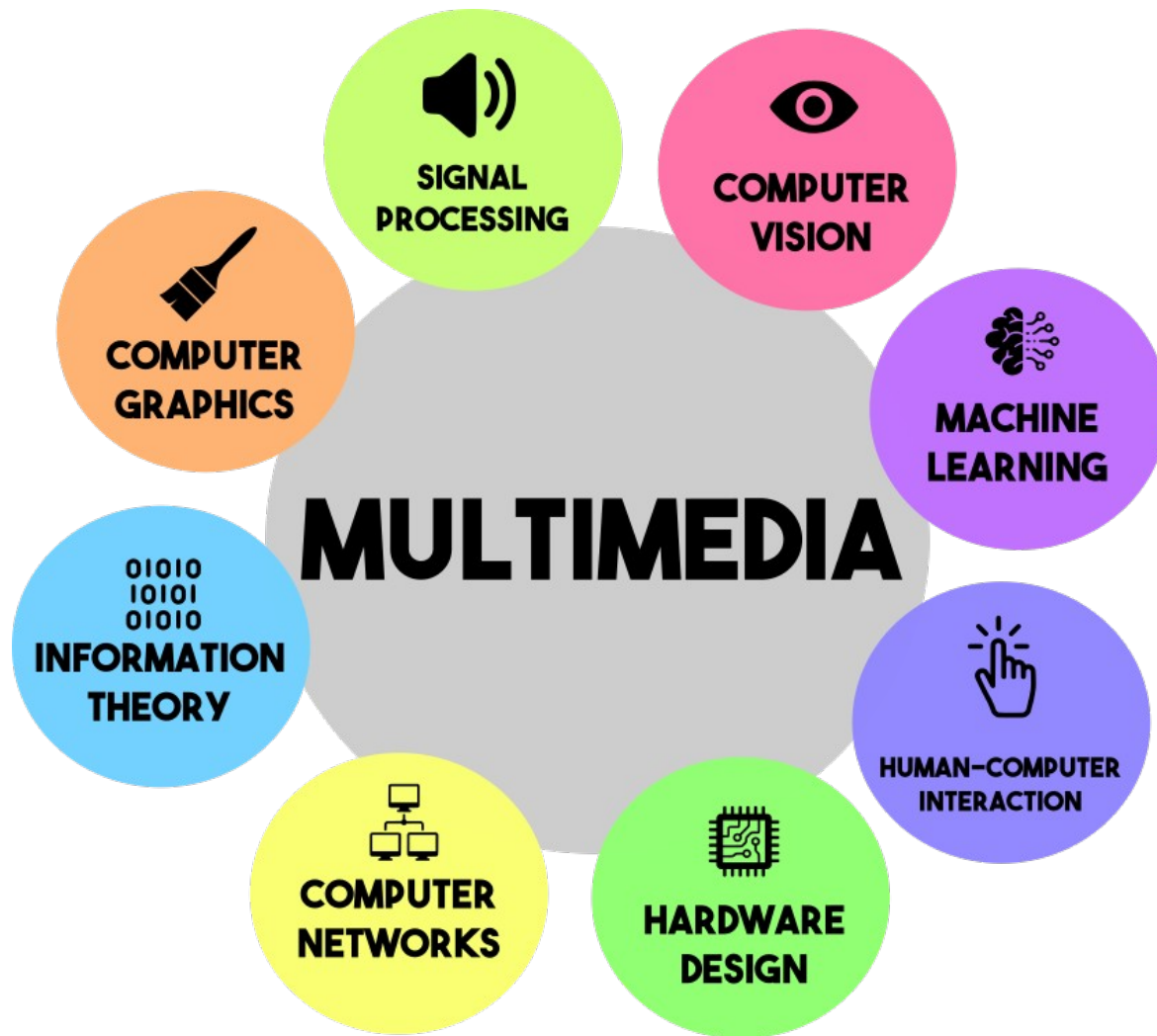
Convergence

Convergence of domains

graphics, visualization, human-computer interaction, computer vision, data compression, computer networks, machine learning, ... are used together.

Convergence of devices

Computers, video players, game consoles, broadcast TV, Internet, converge into a unified multimedia products.



Hypermedia

- Ted Nelson (~1965): HyperText
 - Book: linear medium
 - HyperText: non-linear (interactive)
- Hypermedia: not only text
 - Form of multimedia application
 - WWW – type of hypermedia application

Application domains

- Digital television, video on demand (video + sound)
- Computer games (graphics + sound + interactivity)
- Teleconferences (video + sound)
- Remote lectures (video + sound + slides)
- Telemedicine (video + sound + haptic + manipulation)
- Large databases (e.g. Google, YouTube, Facebook, Amazon, Dropbox)
- Extended reality
- Data visualization (image + sound + interactivity)

Research challenges

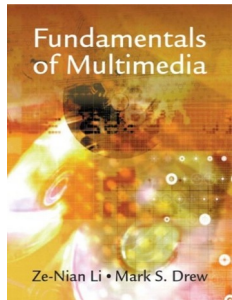
- **Processing and storage**
Content analysis, information retrieval, compression, security, etc.
- **Tools and applications**
Hypermedia systems, content manipulation, user interfaces, multi-modal interaction, content production systems, collaboration systems, etc.
- **Support systems**
Network protocols, quality of service, distribution networks, storage systems, IO devices, etc.

Lectures overview

- Images
- Video
- Sound
- Processing
- Compression
- Retrieval
- Interactivity

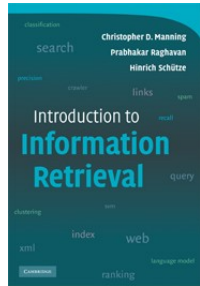
Literature

- Slides + lecture notes available at online Classroom (Učilnica)
- Multimedia overview, general topics



Li Ze-Nian, M. S. Drew,
Fundamentals of Multimedia, 2010.

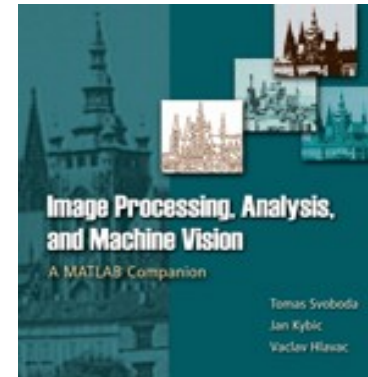
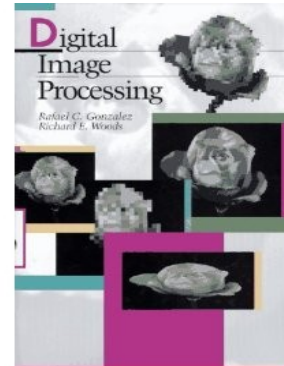
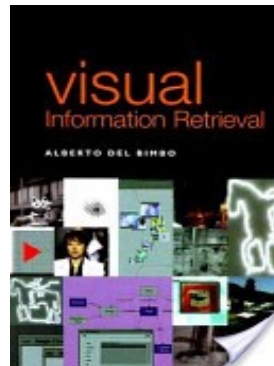
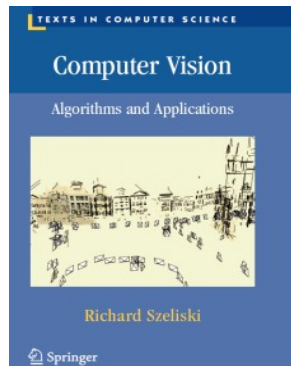
- General information retrieval concepts



C. D. Manning, P. Raghavan, H. Schütze,
[Introduction to Information Retrieval](#), Cambridge University Press. 2008.

Additional literature

- R. Szeliski: Computer Vision: **Algorithms and Applications**
- A. del Bimbo: Visual information retrieval
- Gonzalez and Woods: Digital Image Processing
- Sonka, Hlavac, Boyle: Image Processing, Analysis, and Machine Vision
- J. O. Smith III, **Introduction to Digital Filters**

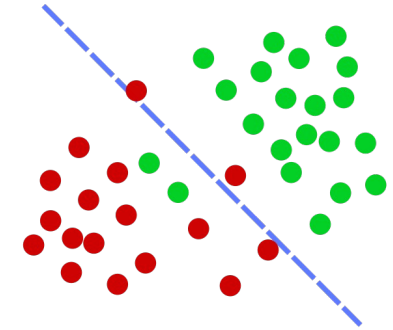


Machine learning for multimedia

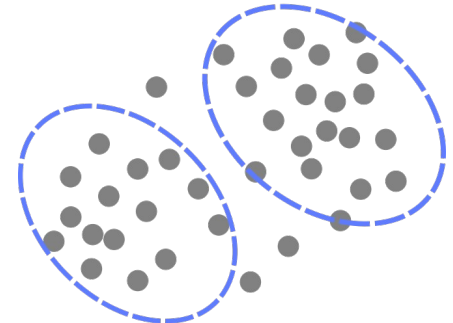
- Machine learning != artificial intelligence
- Model = Function approximation
 - Without explicit programming
 - Improve with data
- Manipulate content
- Extract information

Learning scenarios

- Supervised learning
 - Known output
 - Optimization of objective function
 - Classification, regression
- Unsupervised learning
 - No annotations
 - Knowledge (structure) discovery, data mining
 - Clustering, latent variable estimation
- *Reinforcement learning*



supervised learning



unsupervised learning

Prediction model

- Algorithm with parameters
 - Learnable paramaters
 - Hyperparameters
- Input (sample)
 - Vector of attributes
 - Image, waveform, ...
- Output (prediction)
 - Class
 - Property

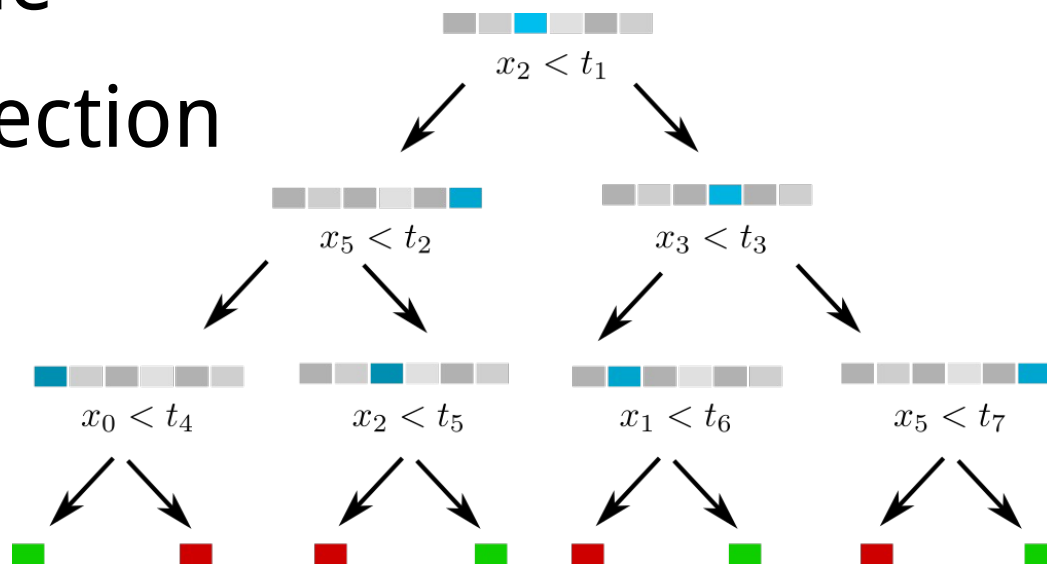


Classification

- Fixed number of classes
 - Binary – yes/no
 - Multi-class
- Use-cases in multimedia
 - Detection (interactivity)
 - Object categorization (retrieval)
 - Tracking

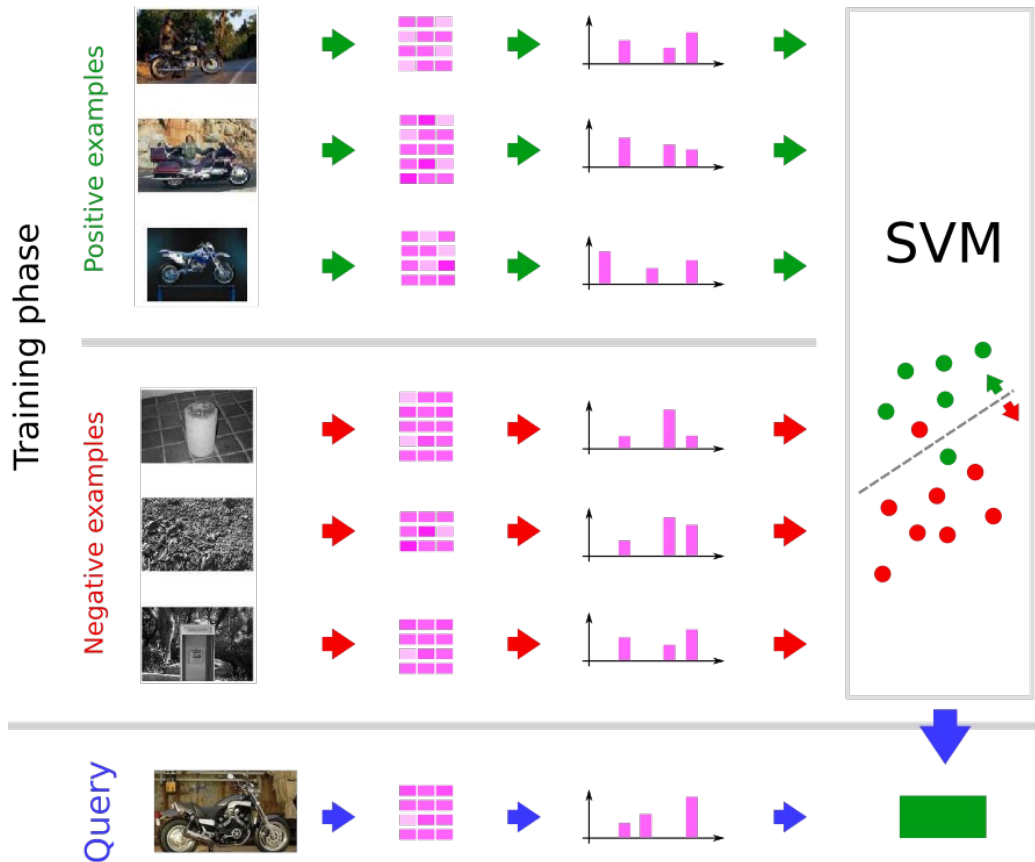
Decision tree

- Inner nodes – test attribute value
- Leaf nodes - outcome
- Greedy attribute selection
- Generalization
 - Boosted trees
 - Random forest



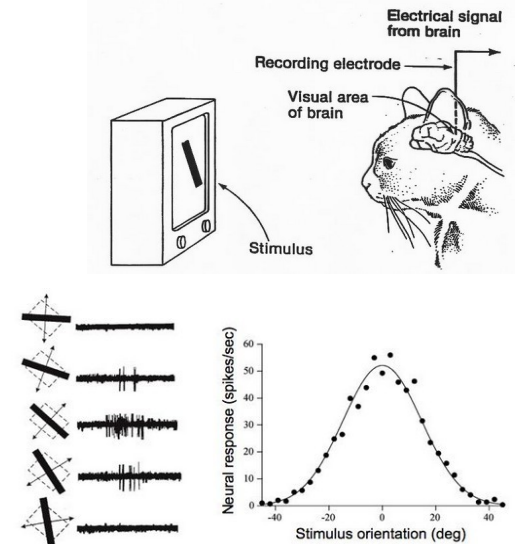
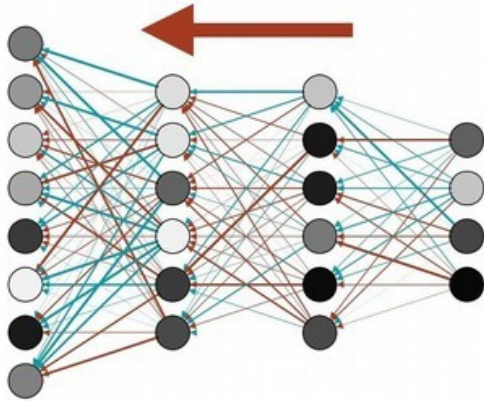
Support vector machine

- Supporting samples
 - Close to boundary
- Kernel space



Artificial neural network

- Biological motivation (~1960)
- Character recognition
- Large number of parameters

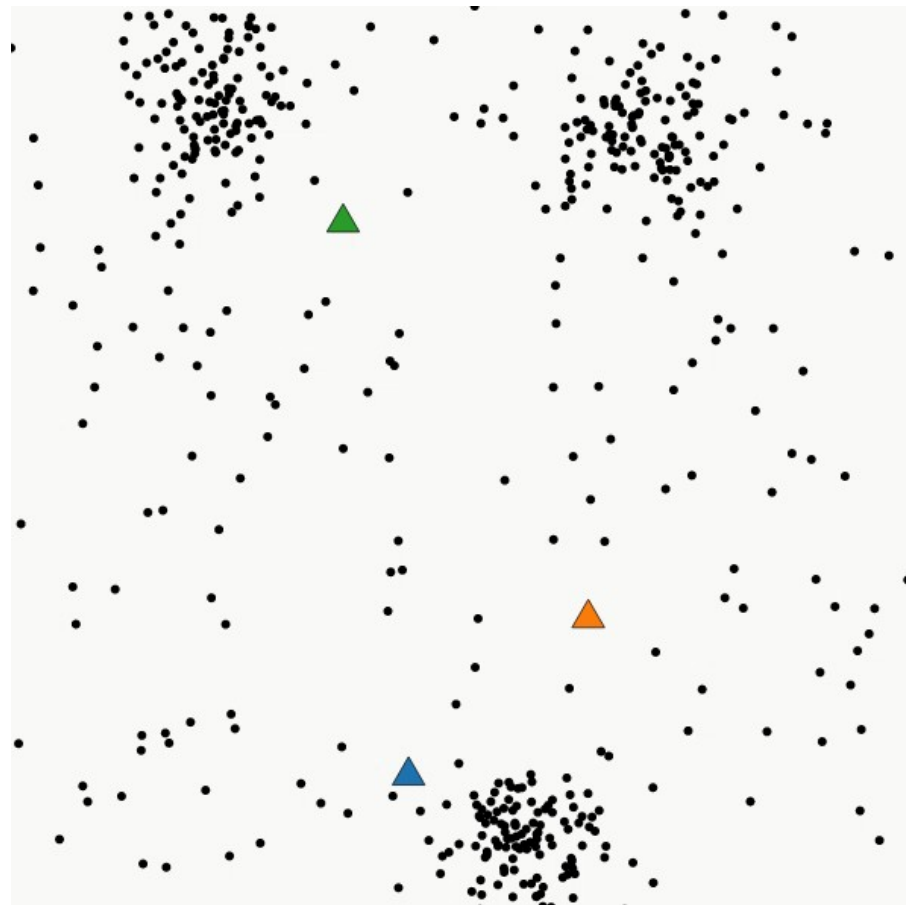


Clustering

- Input – feature vectors
- Output – cluster assignments (labels)
- Chicken / egg problem
- Use cases in multimedia
 - Segmentation
 - Visual dictionary formation
 - Efficient searching

K-means clustering

- Fixed number of clusters (K)
- Randomly initialize cluster centers
- Iterate until convergence:
 - For all p : assign p to cluster k if it is the closest
 - Recompute centers as mean of assigned points



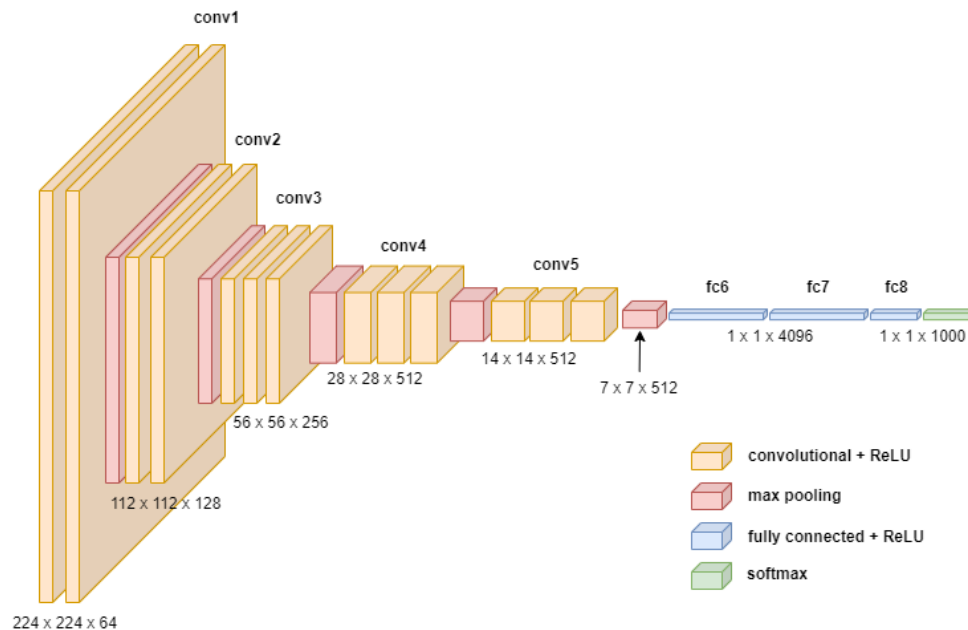
Hierarchical clustering

- Hierarchy of clusters
- Distance metric
- Iterative algorithm
 - Agglomerative
 - Divisive



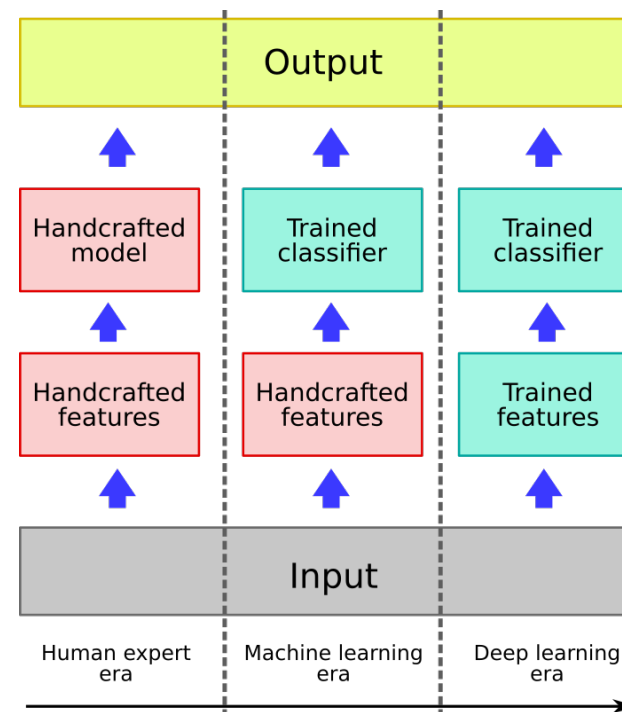
Deep learning

- Large models
 - Neural networks, convolution
 - Many parameters
- Highly non-linear
- Optimization
 - Automatic differentiation
 - Backpropagation of loss function
 - Gradient descent



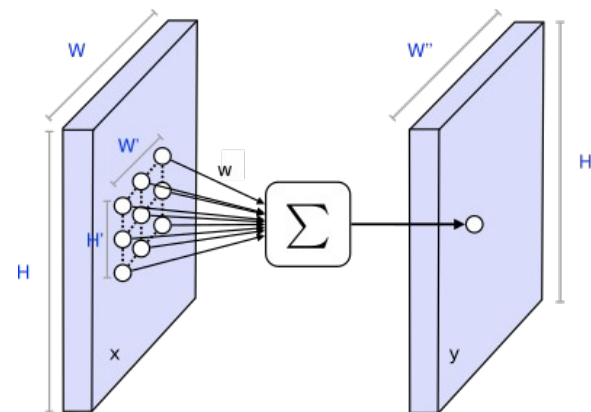
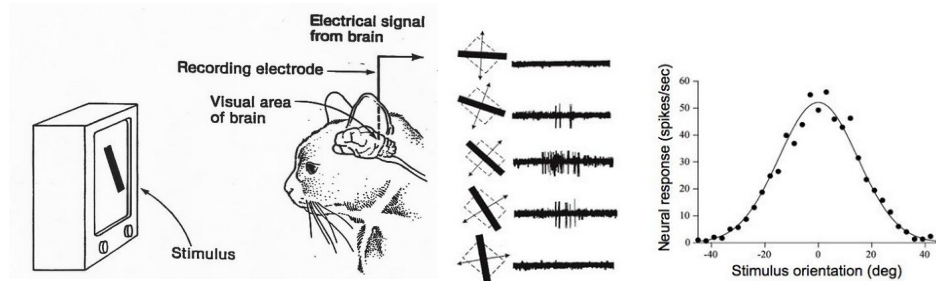
Big data and end-to-end learning

- End-to-end learning
 - Learning features and classifier
 - From pixels to high-level decisions
- Big data
 - More data (Internet, Mechanical Turk)
 - Hardware (storage, GPUs)
 - Learning techniques



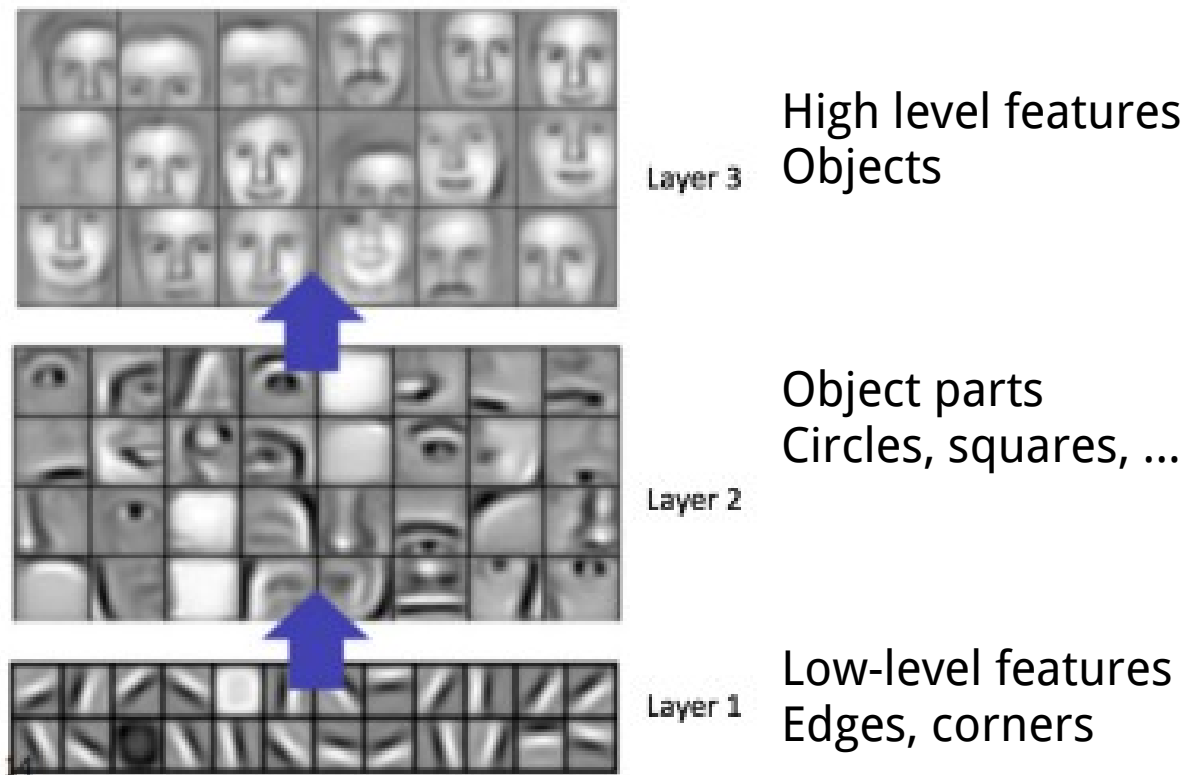
Convolutional neural networks

- Neural networks
 - Biological motivation (~1960)
 - Character recognition
 - High number of parameters
- Convolution
 - Receptive field
 - Same operation on entire image
 - Reduced number of parameters



Hierarchy of filters

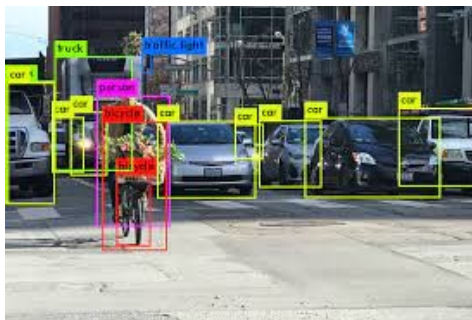
- Training
 - Back-propagation
 - Gradient descent
- Layers
 - Convolutional
 - Fully-connected
 - Max-pooling
 - Soft-max



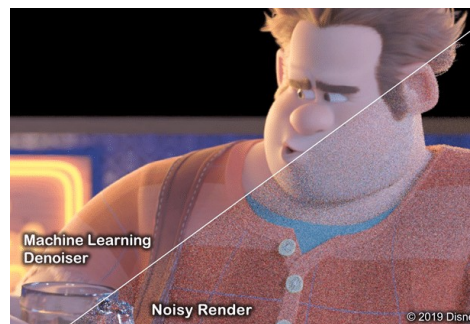
Differentiable programming

- Input

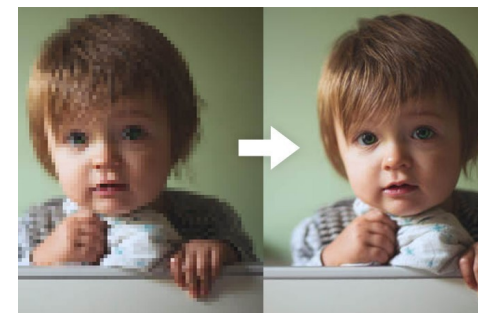
- Image
- Video
- Audio



Detection



De-noising



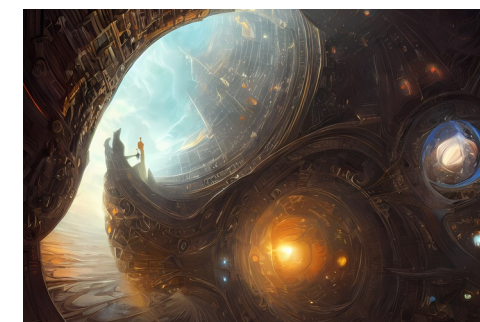
Super resolution

- Output

- Labels
- Regions
- Transformed image
- Generated image
- Generated video



Style transfer



Text-to-image