Multimedia Systems

About the lecturer

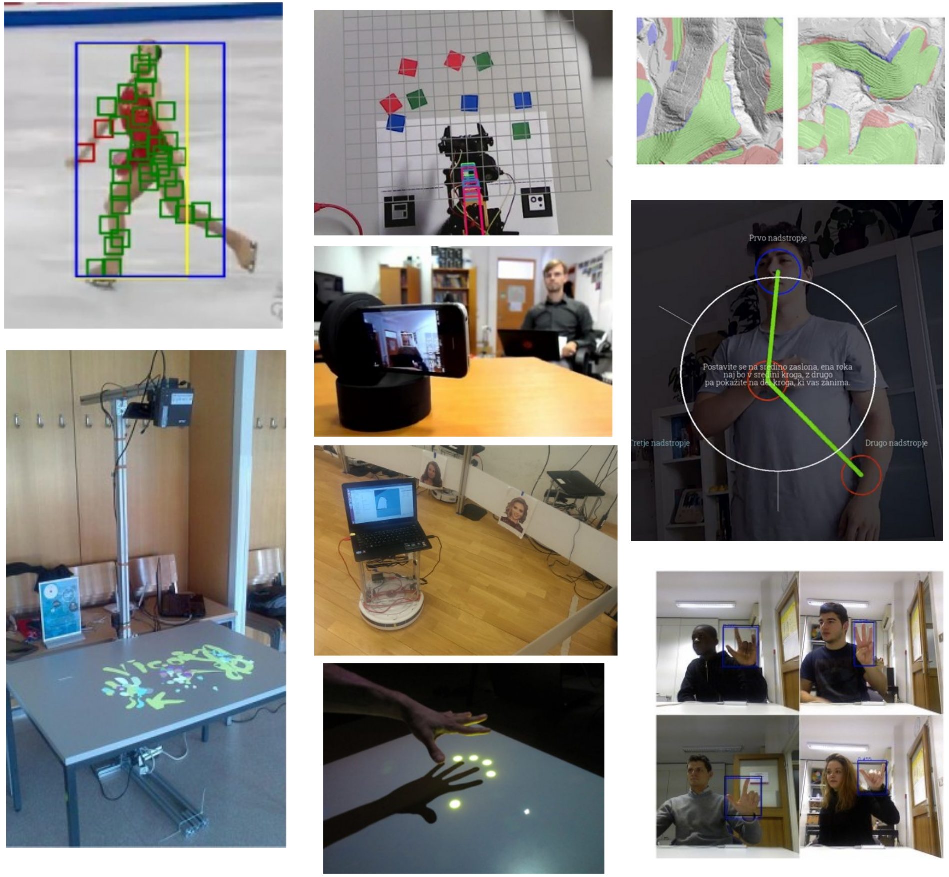
Luka Čehovin Zajc, PhD

Assistant Professor

Visual Cognitive Systems Laboratory

Room R2.39

luka.cehovin@fri.uni-lj.si


Prvo nadstropje
Postavite se na sredino zaslona, ena roka naj bo v sredini kroga, z drugo pa pokažite na del kroga, ki vas zanima.
Tretje nadstropje
Drugo nadstropje

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) -50%**
  - 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

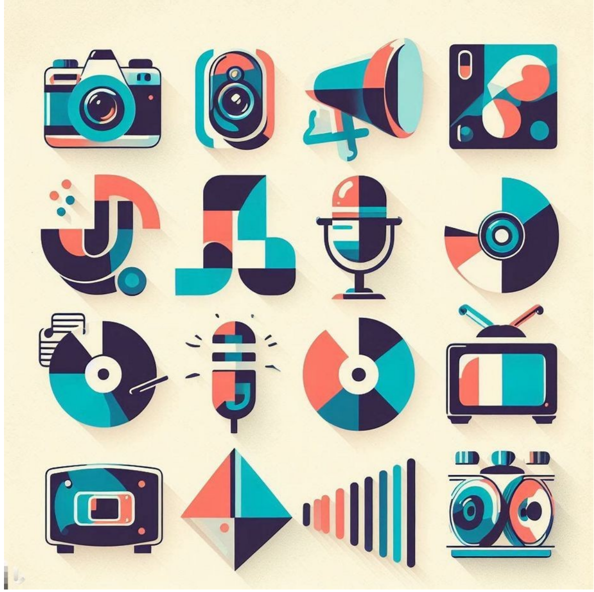
Laboratory exercises

- 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
  - Labs = Presentation + Consultations + Defenses

Project assignment

- Alternative to regular laboratory exercises
- In-depth 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 / multi-touch 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.

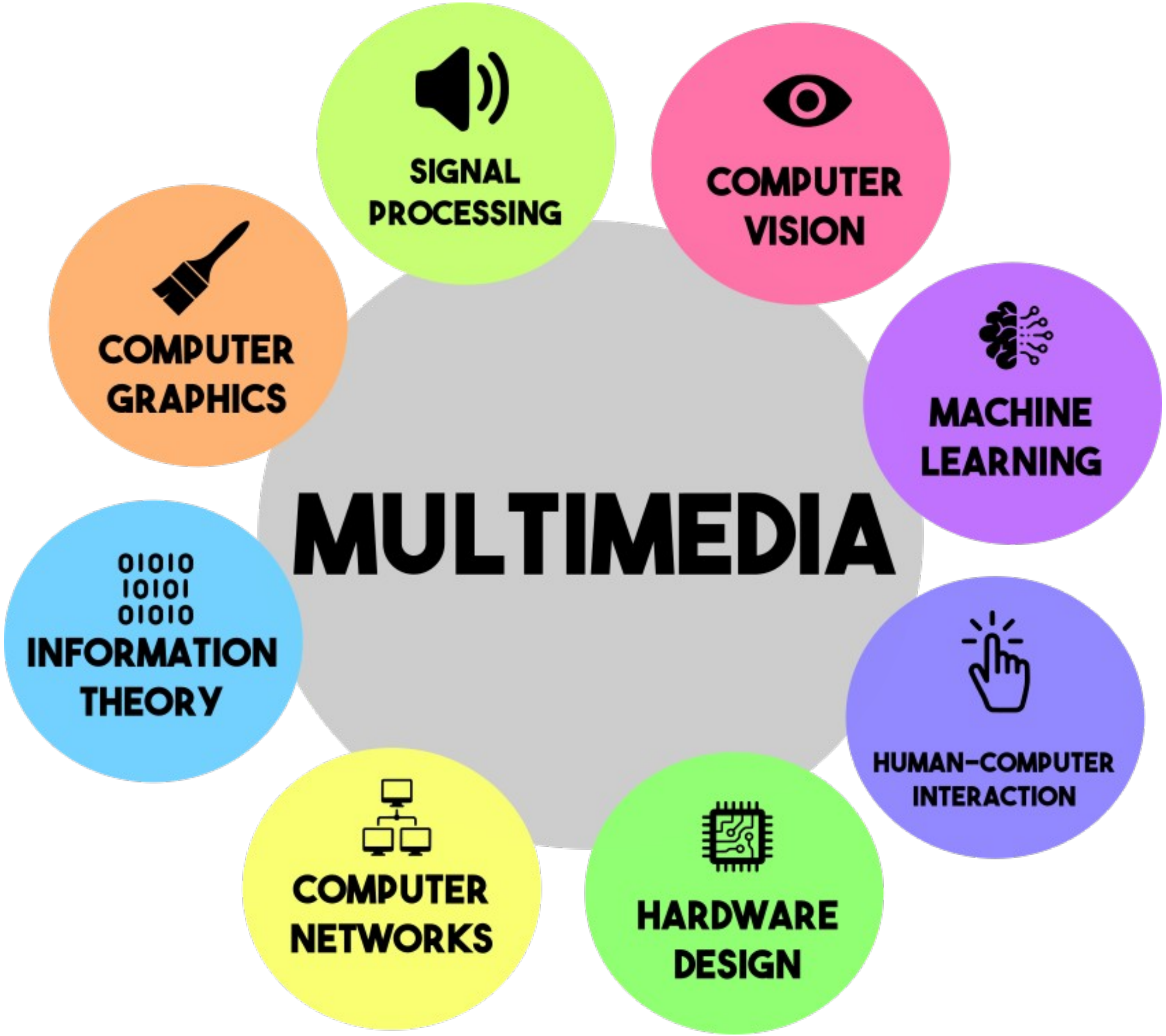
Convergence

Devices

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

Domains

graphics, visualization, human-computer interaction, computer vision, data compression, signal processing, computer networks, machine learning ...



MULTIMEDIA

SIGNAL PROCESSING

COMPUTER VISION

MACHINE LEARNING

HUMAN-COMPUTER INTERACTION

HARDWARE DESIGN

COMPUTER NETWORKS

INFORMATION THEORY

COMPUTER GRAPHICS

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, storage**
  Content analysis, information retrieval, compression, security, etc.
- **Tools, applications, methods**
  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.

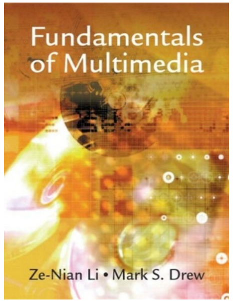
Lectures overview

- Images
- Video
- Sound

- Processing
- Compression
- Retrieval
- Interactivity

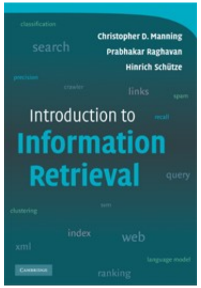
Literature

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



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

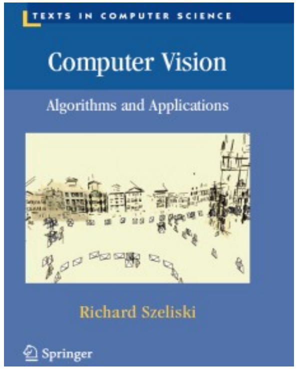
- General information retrieval concepts

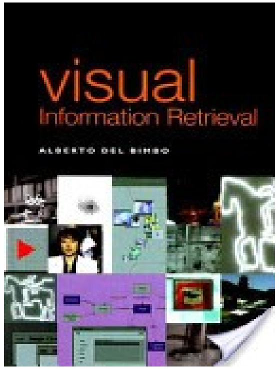


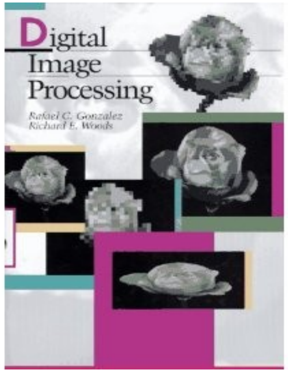
C. D. Manning, P. Raghavan, H. Schütze,
Introduction to Information Retrieval, Cambridge University Press. 2008.

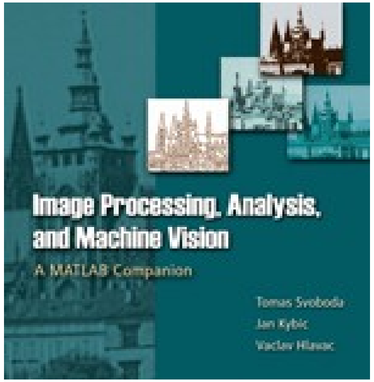
Additional literature

- 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





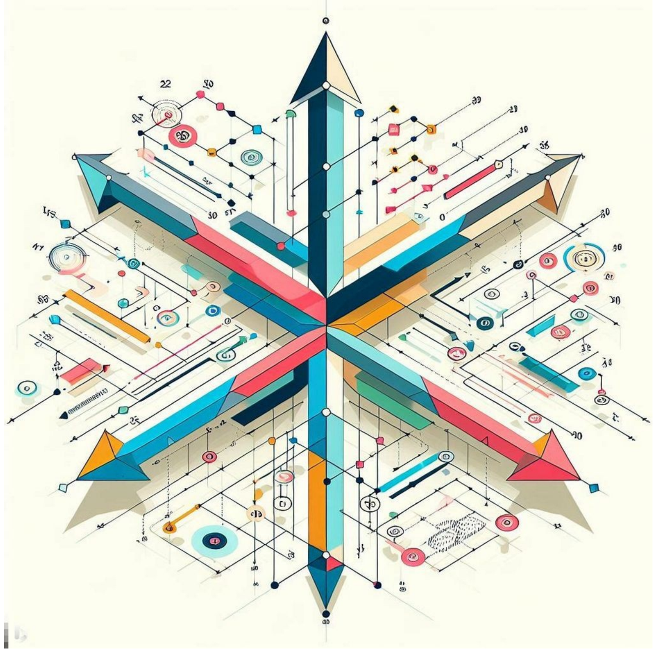




Multimedia and machine learning

Representation

- Vector of values
  - Encoding data properties
  - Embedding – special case (structured space)
- Task dependent
  - Select the right representation
  - General vs. specialized
  - What to describe?



Examples of representations

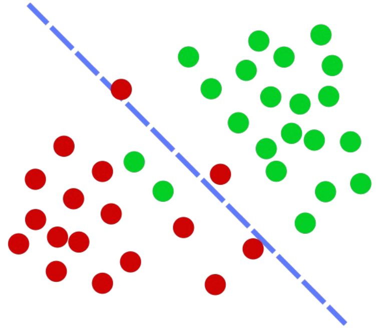
- Image / video
  - Sequence of pixels
  - Histogram, average color
  - Bag of words vector
  - Frequency space
  - RLE
  - Image embedding
- Audio
  - Waveform
  - Spectrogram
  - Vocoder parameters
  - Audio embedding
- Text
  - Word frequency
  - Text embedding

Machine learning

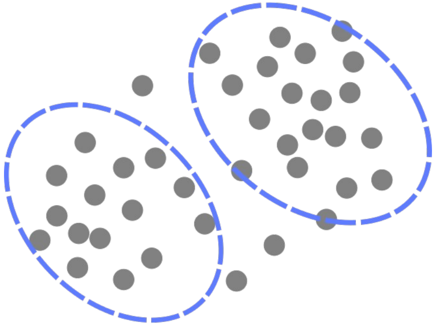
- Machine learning != artificial intelligence
- Model = Function approximation
  - Without explicit programming
  - Based on given data
  - Improve with more data
- Manipulate content
- Extract information

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*



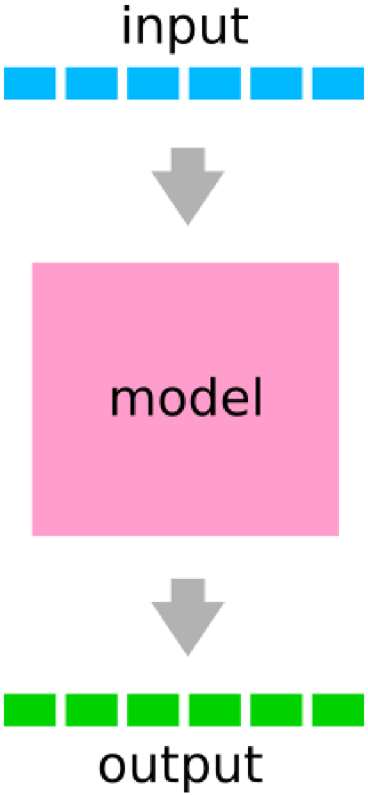
supervised learning



unsupervised learning

Prediction model

- Algorithm with parameters
  - Learnable parameters
  - Hyper-parameters
- Input (sample)
  - Vector representation
  - Image, waveform, ...
- Output (prediction)
  - Class
  - Property



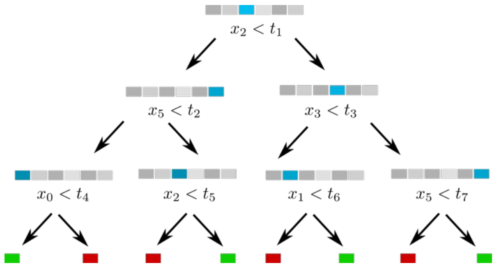
Classification

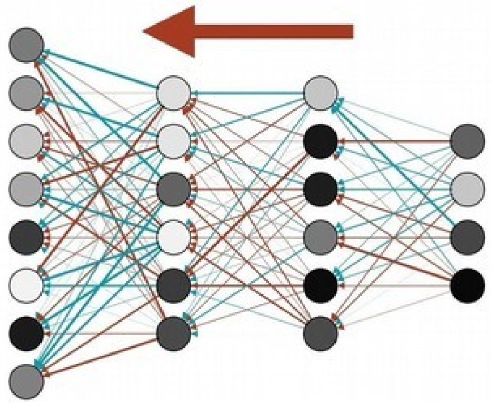
- Fixed number of classes
  - Binary – yes/no
  - Multi-class

- Use-cases in multimedia
  - Object detection, segmentation
  - Object categorization
  - Semantic description

Classification methods

- K nearest neighbors
- Decision trees
- Random forest
- Boosting
- Support vector machines
- Artificial neural networks





Clustering

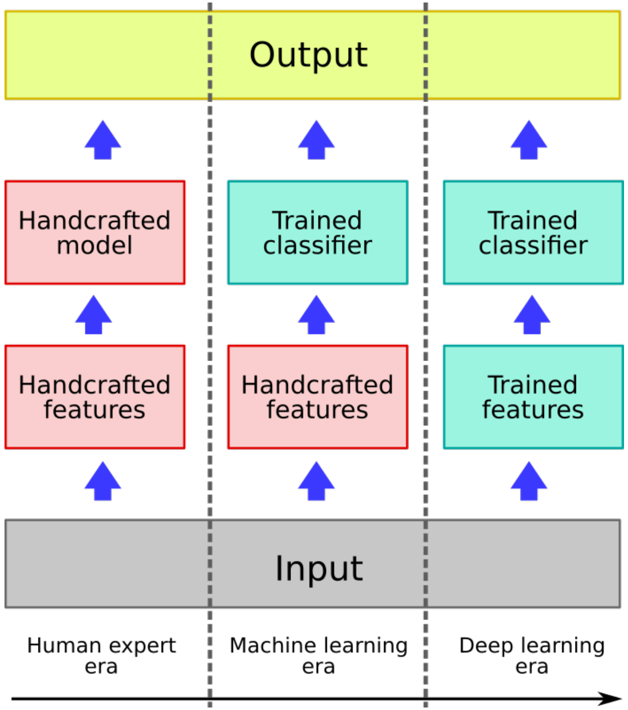
- Input – feature vectors
- Output – cluster assignments (labels)
- Chicken / egg problem
- Use cases in multimedia
  - Segmentation – grouping pixels
  - Visual dictionary formation – grouping descriptors
  - Efficient searching – grouping representations

Clustering methods

- Hierarchical
- K-means, mean-shift
- Spectral
- Message passing

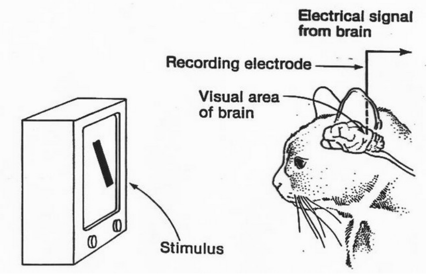
Big data and end-to-end learning

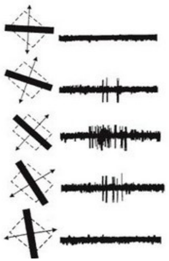
- End-to-end learning
  - Learning features and classifier
  - From pixels to high-level decisions
- Big data
  - More data (Internet, Mechanical Turk)
  - Hardware (storage, GPUs)
  - Learning techniques

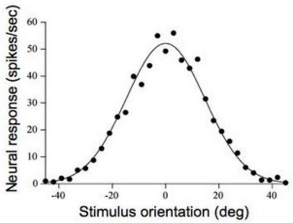


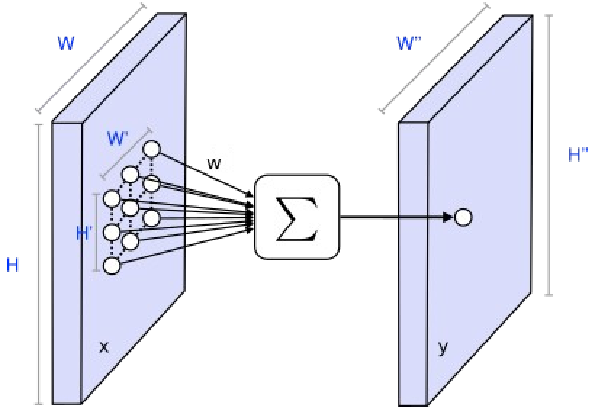
Convolutional neural networks

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



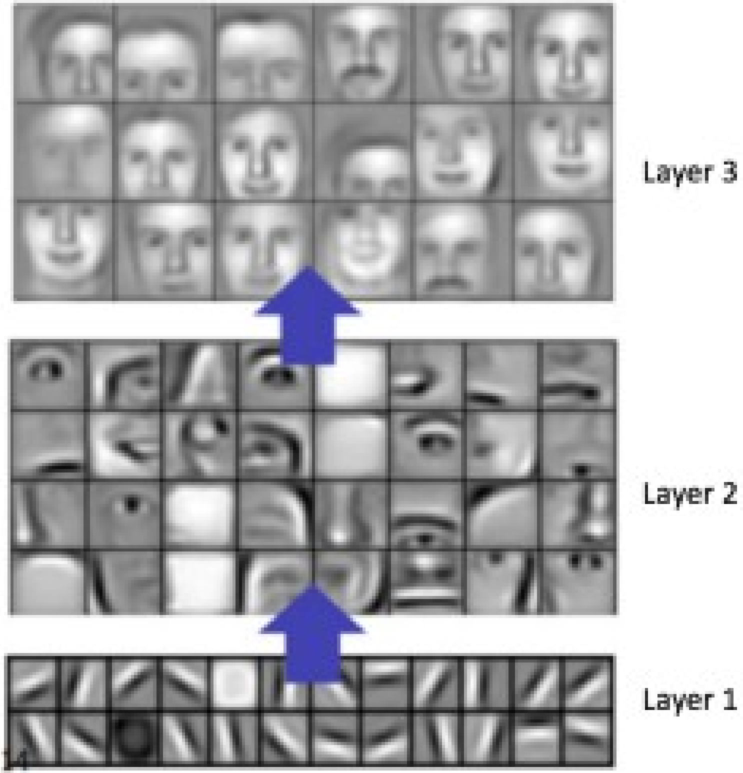






Hierarchy of operations

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



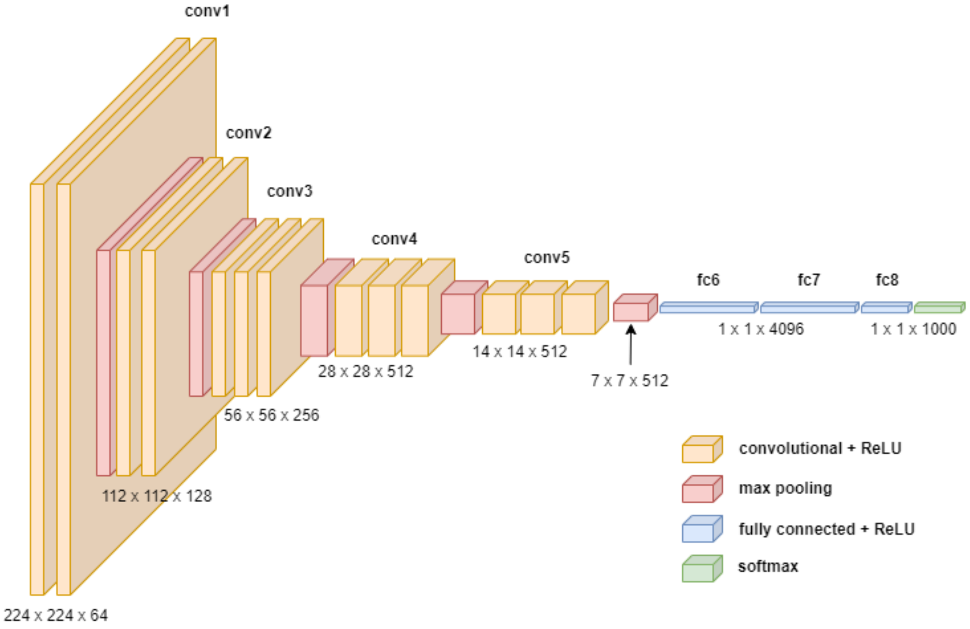
High level features
Objects

Object parts
Circles, squares, ...

Low-level features
Edges, corners

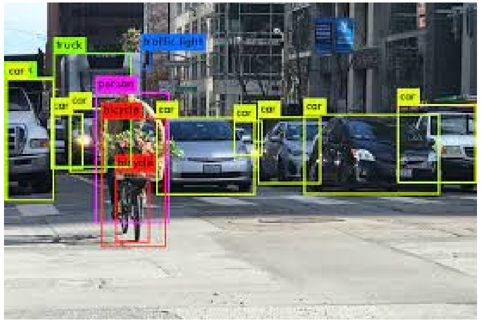
Deep learning

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

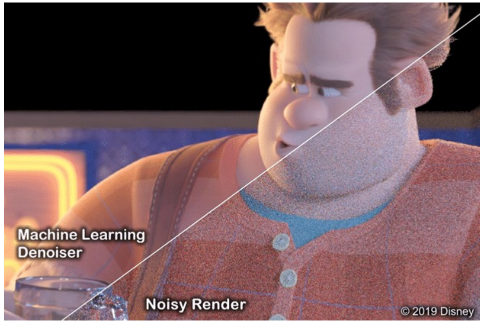


Differentiable programming

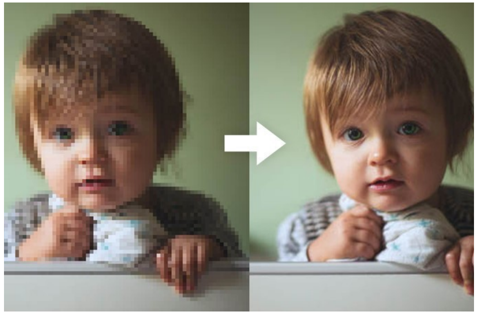
- Input
  - Image
  - Video
  - Audio
- Output
  - Labels
  - Regions
  - Transformed image
  - Generated image
  - Generated video



Detection



De-noising



Super resolution

Style transfer

Text-to-image