Machine Learning & Object Recognition
2016 - 2017

Cordelia Schmid

Jakob Verbeek

Content of the course

- Visual object recognition
- Machine learning

Practical matters

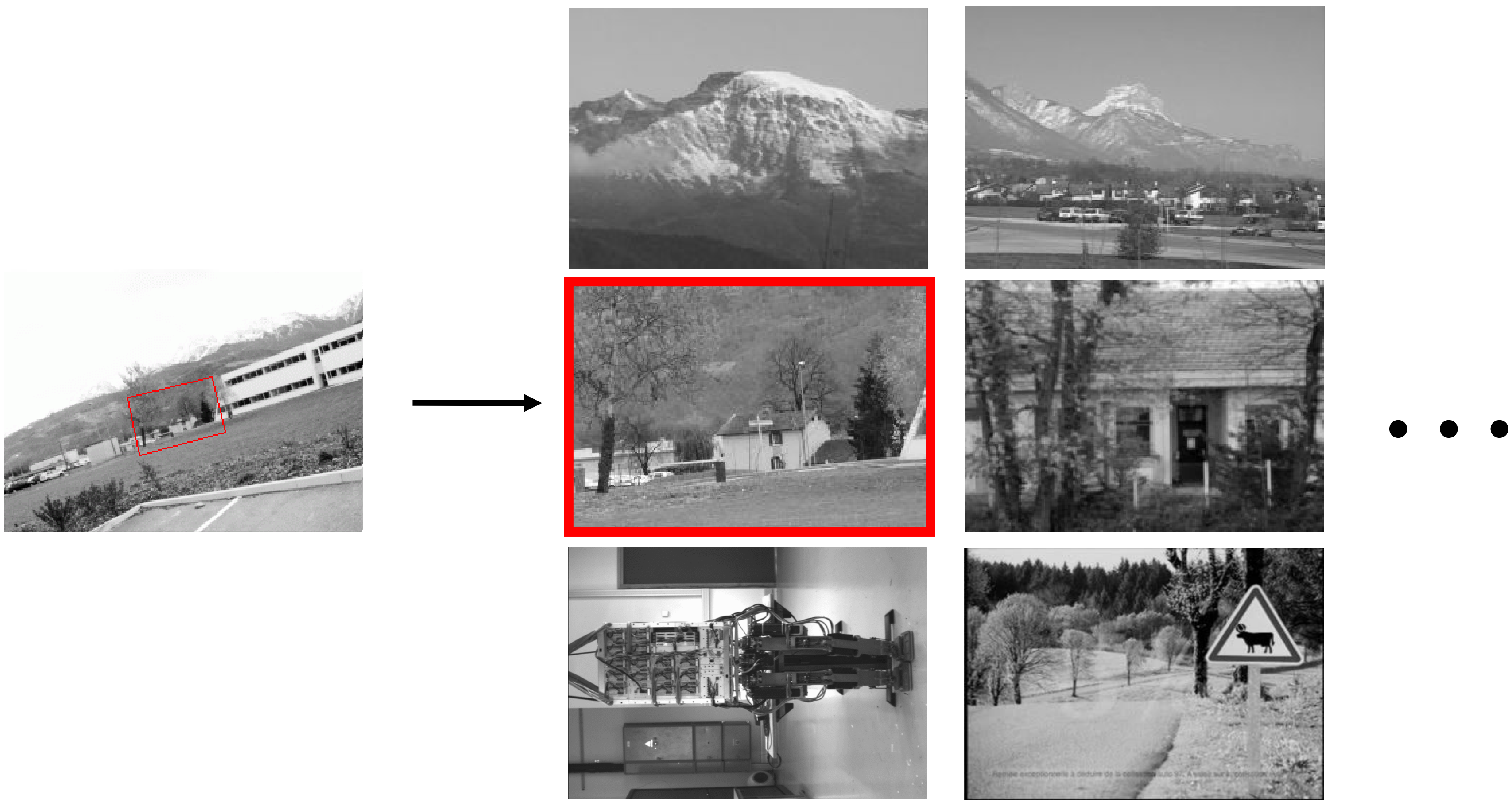
- Online course information
  - Schedule, slides, papers
  - http://thoth.inrialpes.fr/~verbeek/MLOR.16.17.php

- Grading: Final grades are determined as follows
  - 50% written exam,
  - 25% paper presentation,
  - 25% quizes on the presented papers

- Paper presentations:
  - each student presents once
  - each paper is presented by two students
  - presentations last for 15~20 minutes, time yours in advance!

Visual recognition - Objectives

- Retrieval of **particular** objects and scenes
- Accuracy and scalability to large databases



Visual object recognition - Objectives

- Detection of object **categories**
  - is there a … in this picture
- More generally: relevance of labels (action, place, ...)

person

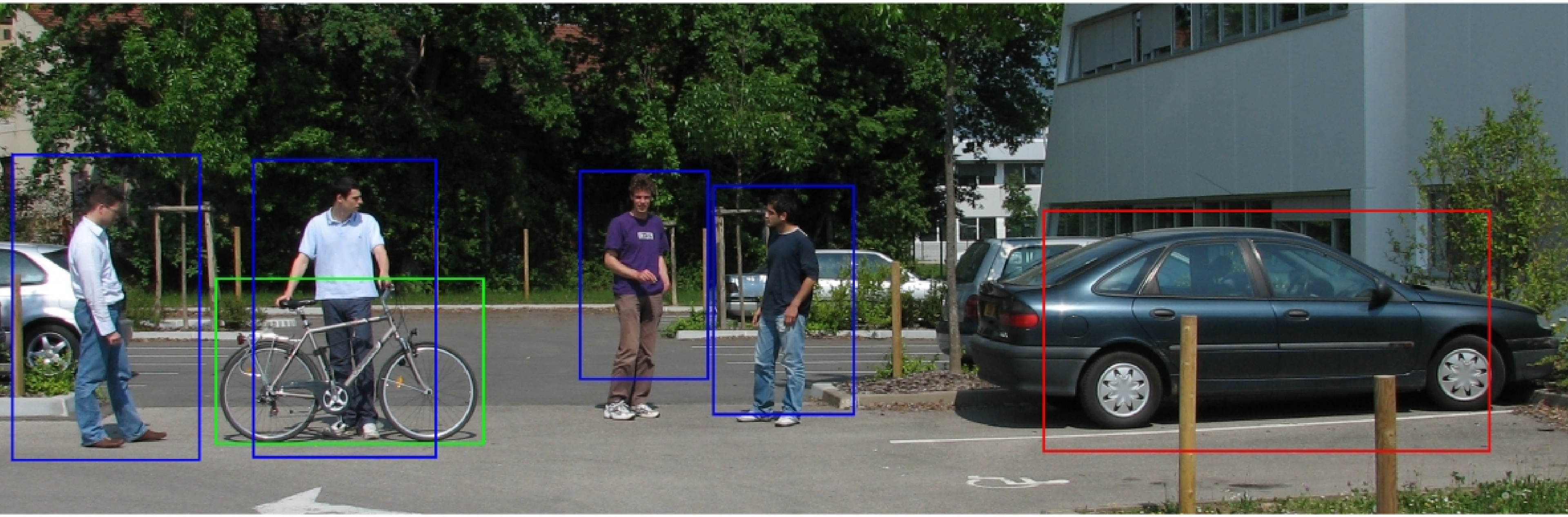
glass

drinking

indoors

Visual recognition - Objectives

- **Localization** of object categories
  - where are the … in this image
- Predict bounding boxes around category instances



Visual recognition - Objectives

- **Semantic segmentation** of (object) categories
  - Which pixels correspond to ….
- Possibly identifying different category instances

Visual recognition - Objectives

- Human pose estimation
- Self-occlusion and clutter

Visual recognition - Objectives

- Human action recognition in video
- Interaction of people and objects, temporal dynamics

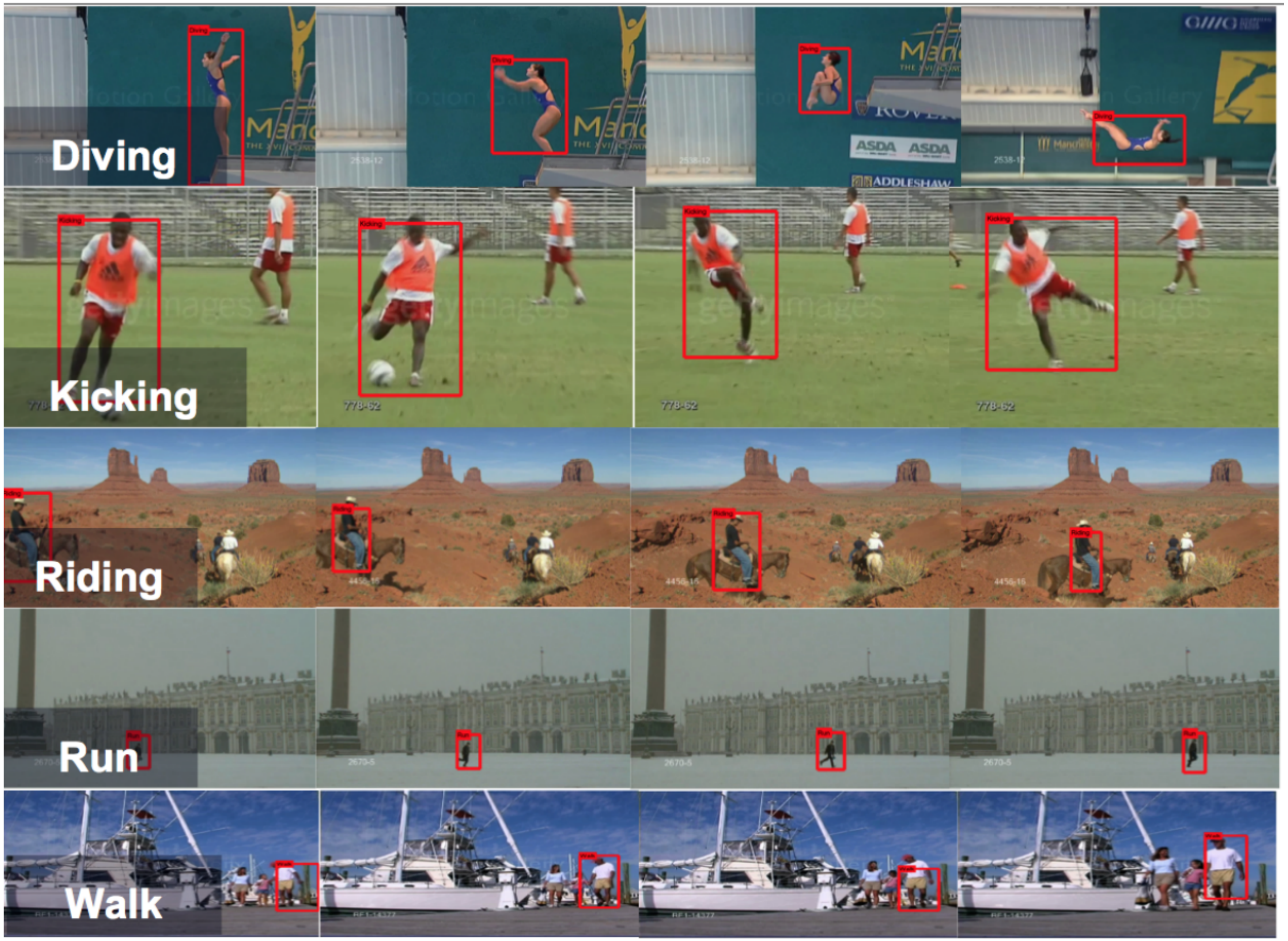
(a) answer-phone (a) get-out-car (a) fight-person (b) push-up (b) cartwheel (b) sword-exercise

(c) high-jump (c) spring-board (c) vault (d) hand-shake (d) high-five (d) kiss

(e) horse-race (e) playing-guitar (e) ski-jet (f) haircut (f) archery (f) ice-dancing

Visual recognition - Objectives

- Human action action localization in time, or space-time



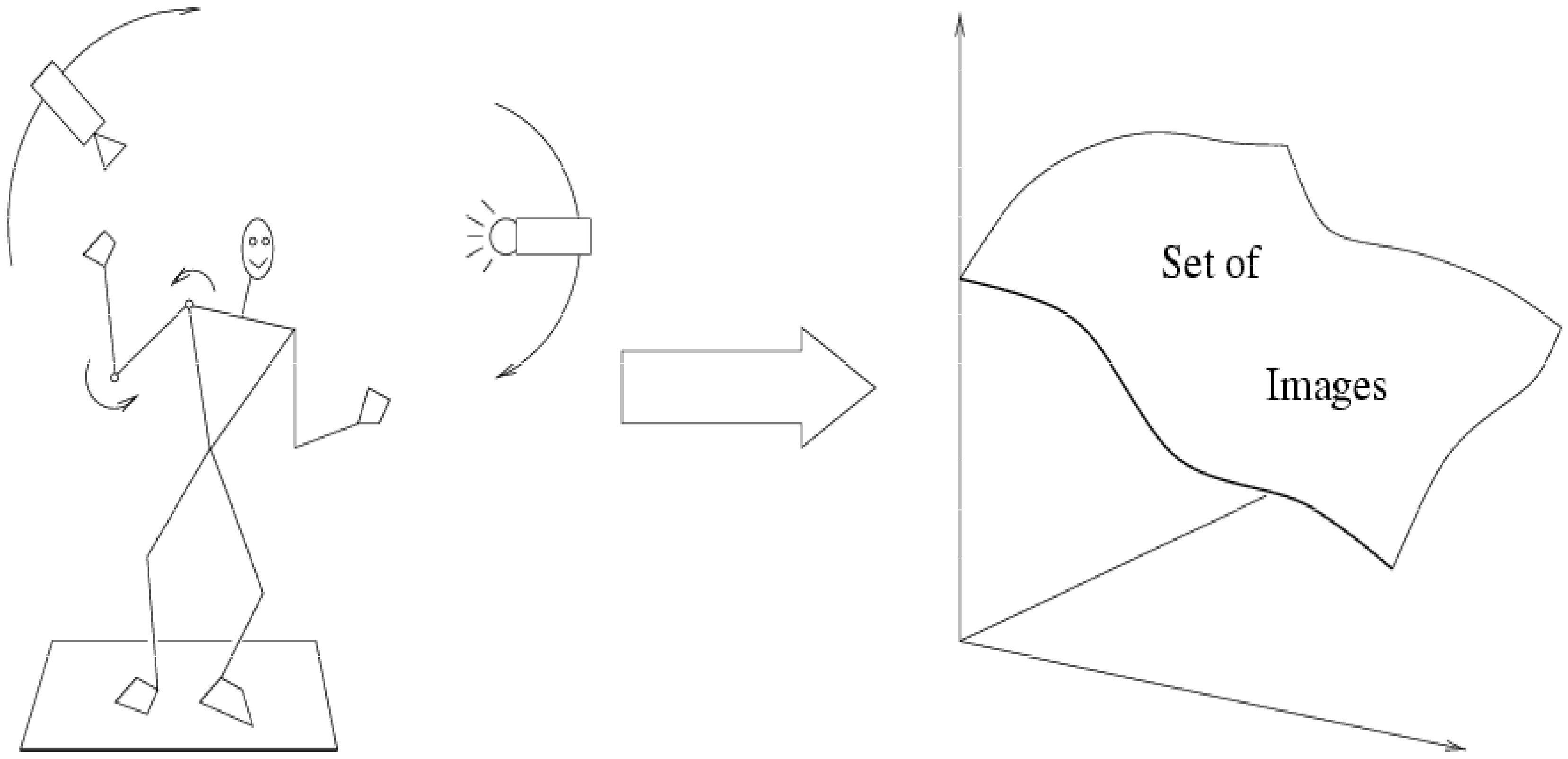
Visual recognition - Objectives

- Image captioning: Given an image produce a natural language sentence description of the image content

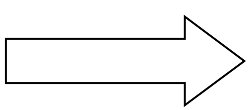
a man and a woman sit on a bench
Prob: 0.0000892

a black and white dog is running through the grass
Prob: 0.00170

Difficulties: within object variations



Variability: Camera position, Illumination,Internal parameters



Within-object variations

Difficulties: within-class variations



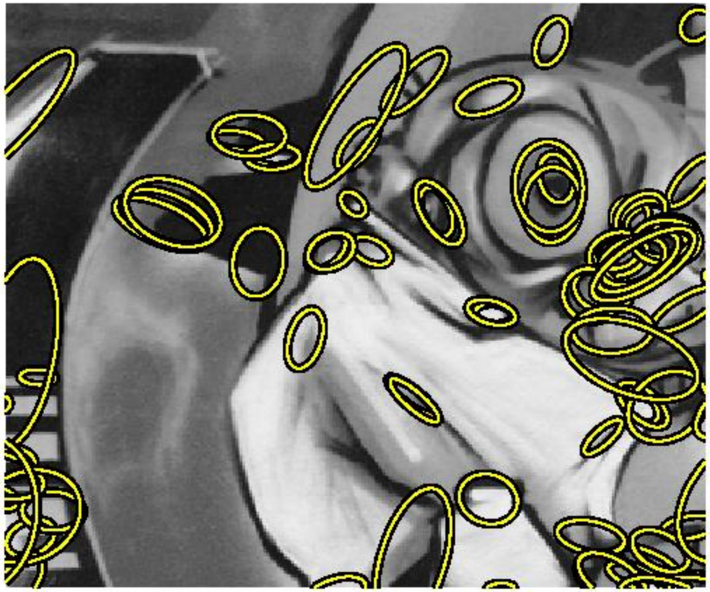



Visual recognition pipeline

- Low-level: Robust image description
  - Appropriate descriptors for objects and categories
  - Possibly unsupervised learning (PCA, clustering, ...)
- High-level: Statistical modeling and machine learning
  - Map low-level descriptors to high-level interpretations
  - Capture the visual variability of specific objects or scenes, but more importantly at the category level
- Today this distinction is less true
  - Learned low-level features
  - Training of low-level and high-level models unified
  - “Deep learning” framework

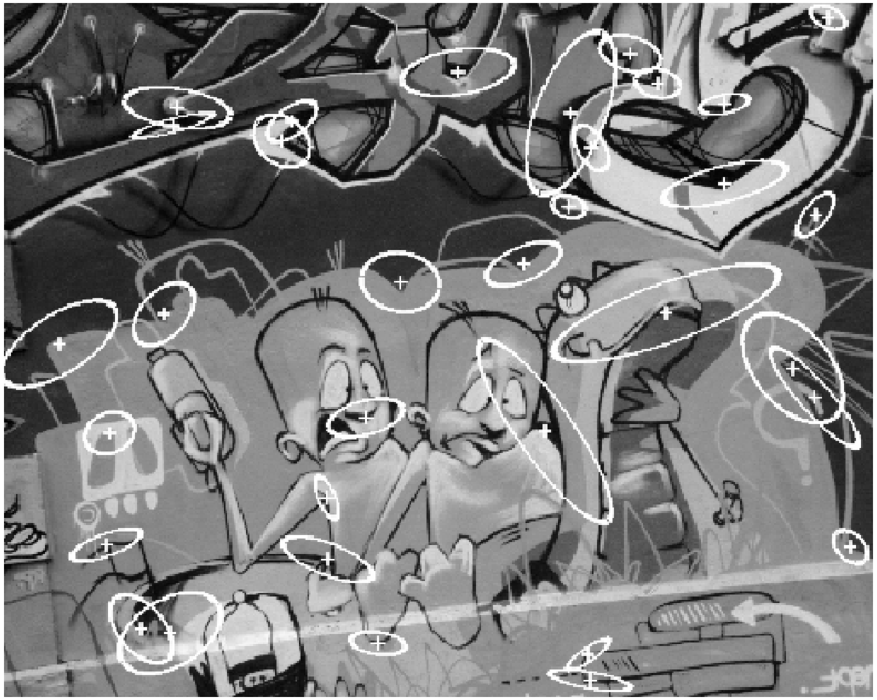
Robust image description

- Scale and affine-invariant keypoint detectors
- Robust keypoint descriptors



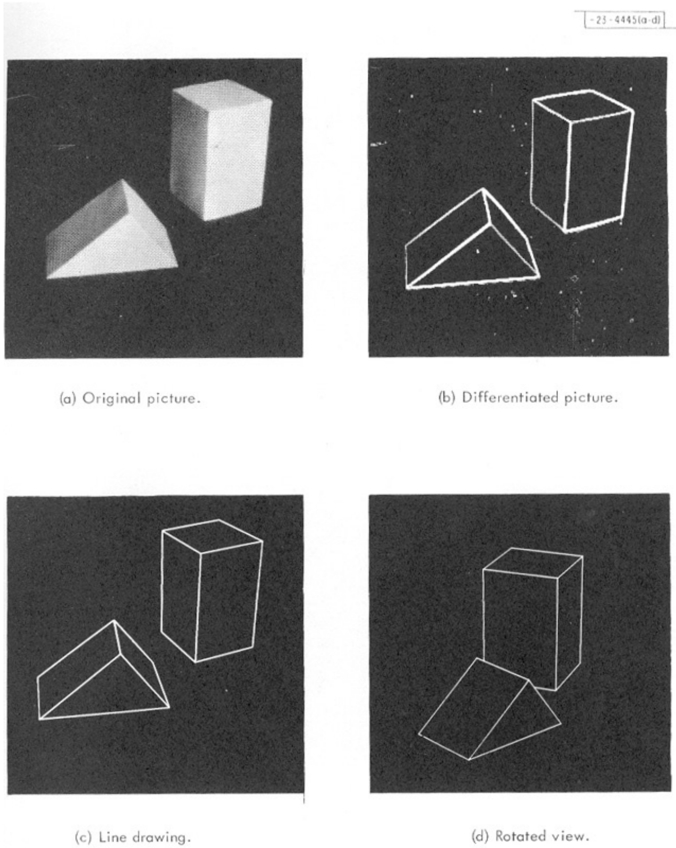
Robust image description

- Matching despite significant viewpoint changes



Why machine learning?

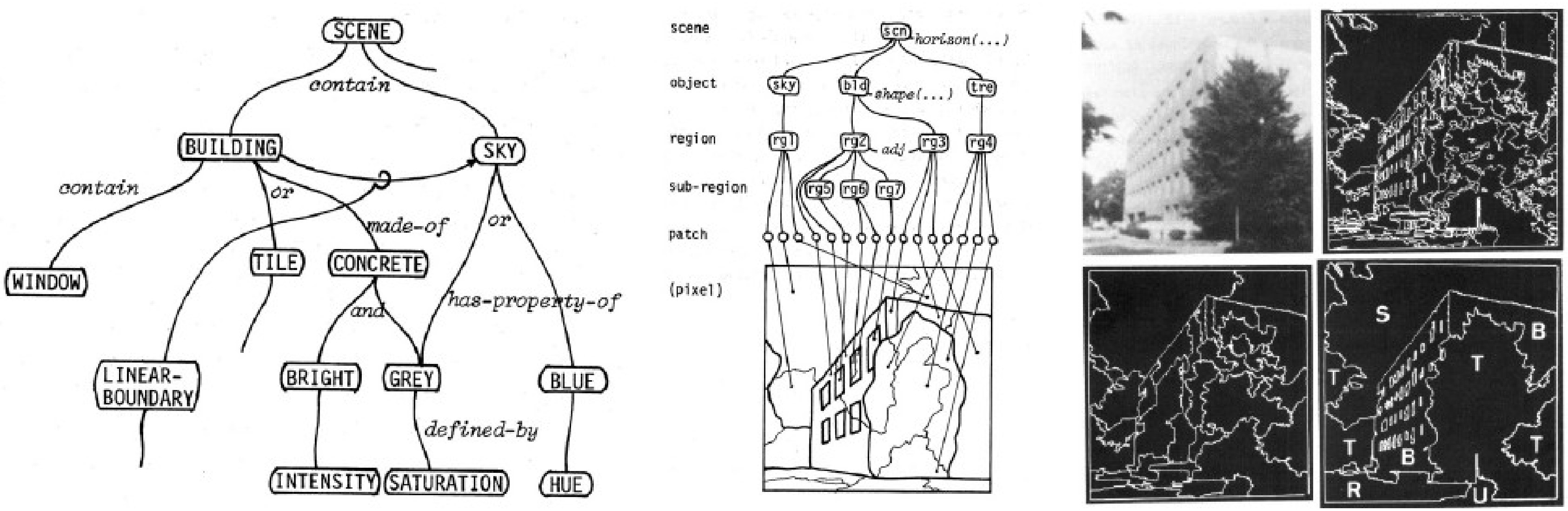
- Early approaches: simple features + handcrafted models
- Can handle only few images, simple tasks



L. G. Roberts, *Machine Perception of Three Dimensional Solids*, Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

Why machine learning?

- Early approaches: manual programming of rules
- Tedious, limited and not directly data-driven



(a) Bottom-up process (b) Top-down process (c) Result

Figure 3. A system developed in 1978 by Ohta, Kanade and Sakai [33, 32] for knowledge-based interpretation of outdoor natural scenes. The system is able to label an image (c) into semantic classes: S-sky, T-tree, R-road, B-building, U-unknown.

Y. Ohta, T. Kanade, and T. Sakai, "An Analysis System for Scenes Containing objects with Substructures," *International Joint Conference on Pattern Recognition*, 1978.

Why machine learning?

- Today: Lots of data, complex tasks
- Instead of trying to encode rules directly, learn them from examples of inputs and desired outputs

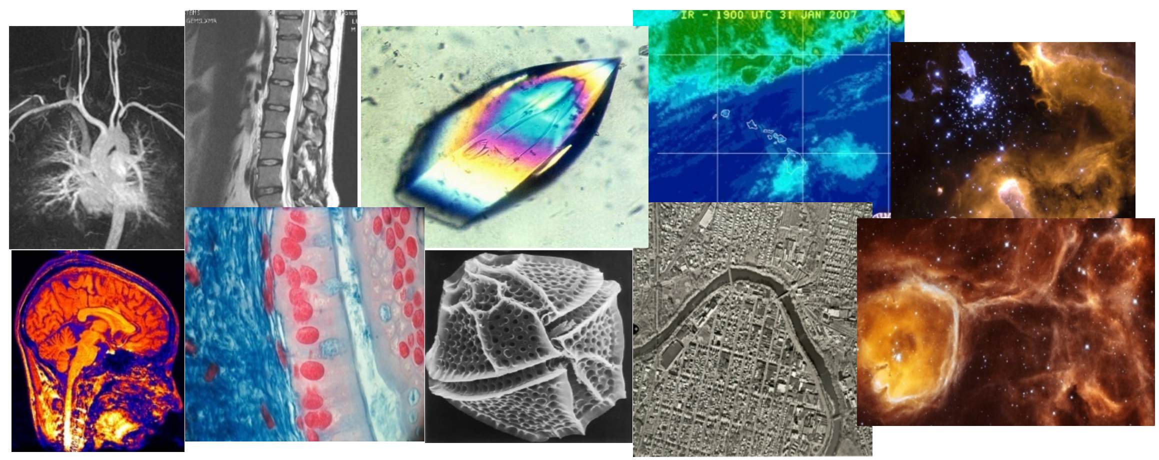
Internet images,
personal photo albums

Movies, news, sports

Why machine learning?

- Today: Lots of data, complex tasks
- Instead of trying to encode rules directly, learn them from examples of inputs and desired outputs

Surveillance and security



Medical and scientific images

Types of learning problems

- Supervised
  - Classification
  - Regression
- Unsupervised
  - Clustering
  - Generative models
- Semi-supervised
- Active learning
- ….

Supervised learning

- Given training examples of inputs and corresponding outputs, produce the “correct” outputs for new inputs

- Two important classic cases:

  – **Classification:** outputs are discrete variables (category labels). Learn a decision boundary that separates one class from the other (separate images with and without cars in them)

  – **Regression:** also known as “curve fitting” or “function approximation.” Learn a continuous input-output mapping from examples (estimate the human pose parameters given an image)

Image captioning

- Given an image produce a natural language sentence description of the image content
- Also supervised learning, but with complex output space

a man and a woman sit on a bench
Prob: 0.0000892

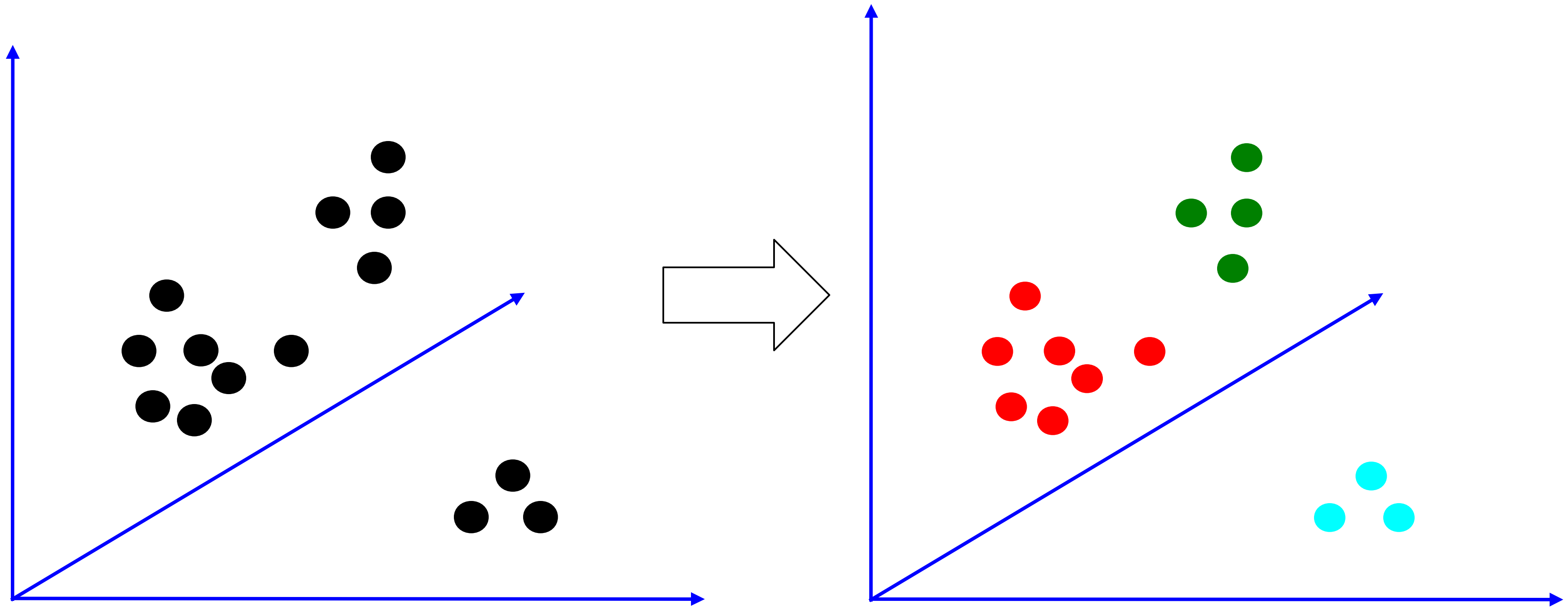
a black and white dog is running through the grass
Prob: 0.00170

Unsupervised Learning

- Given only *unlabeled* data as input, learn some sort of structure from the data
  - Clusters
  - Low-dimensional subspace
- The objective function is typically based on a ``reconstruction'': how well can the original data be explained by the recovered structure?
- Most methods can be (re)formulated as a generative model: fit a model p(x) to ``predict'' data samples
  - Density estimation

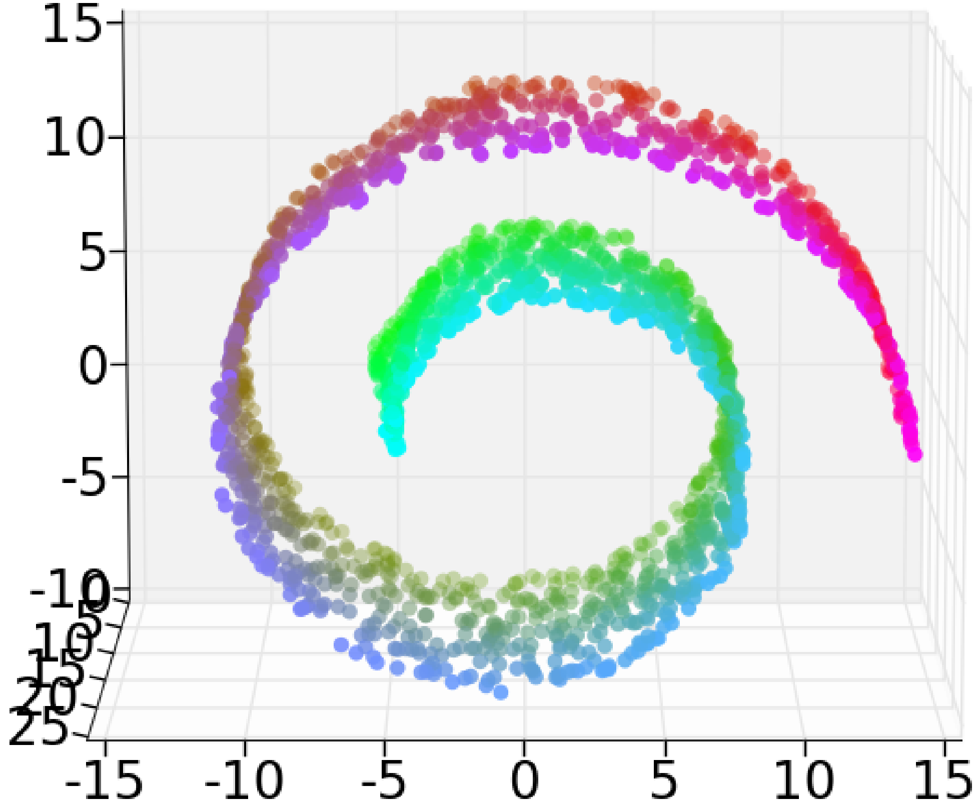
Unsupervised Learning

- Clustering: Discover groups of “similar” data points



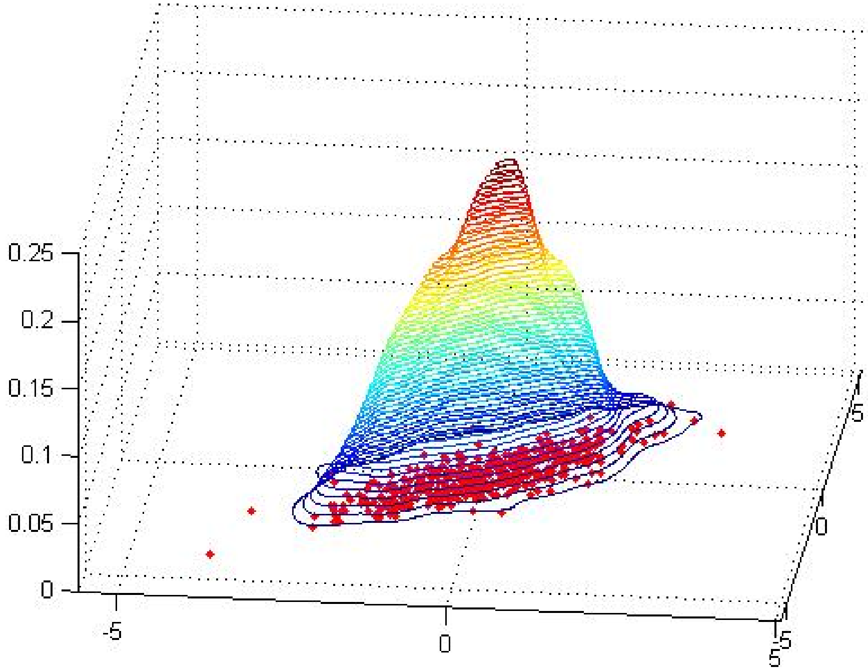
Unsupervised Learning

- **Dimensionality reduction, manifold learning**
  - Discover a lower-dimensional surface on which the data lives



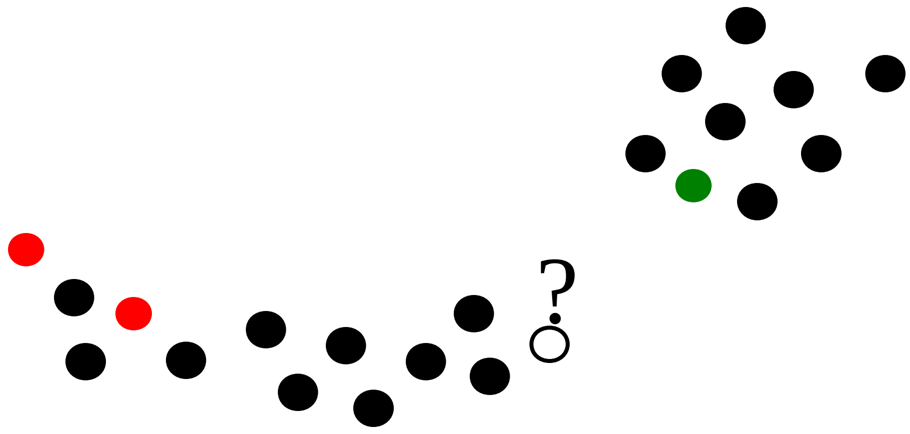
Unsupervised Learning

- **Density estimation**
  - Find a function that approximates the probability density of the data (i.e., value of the function is high for “typical” points and low for “atypical” points)
  - Can be used for **anomaly detection**



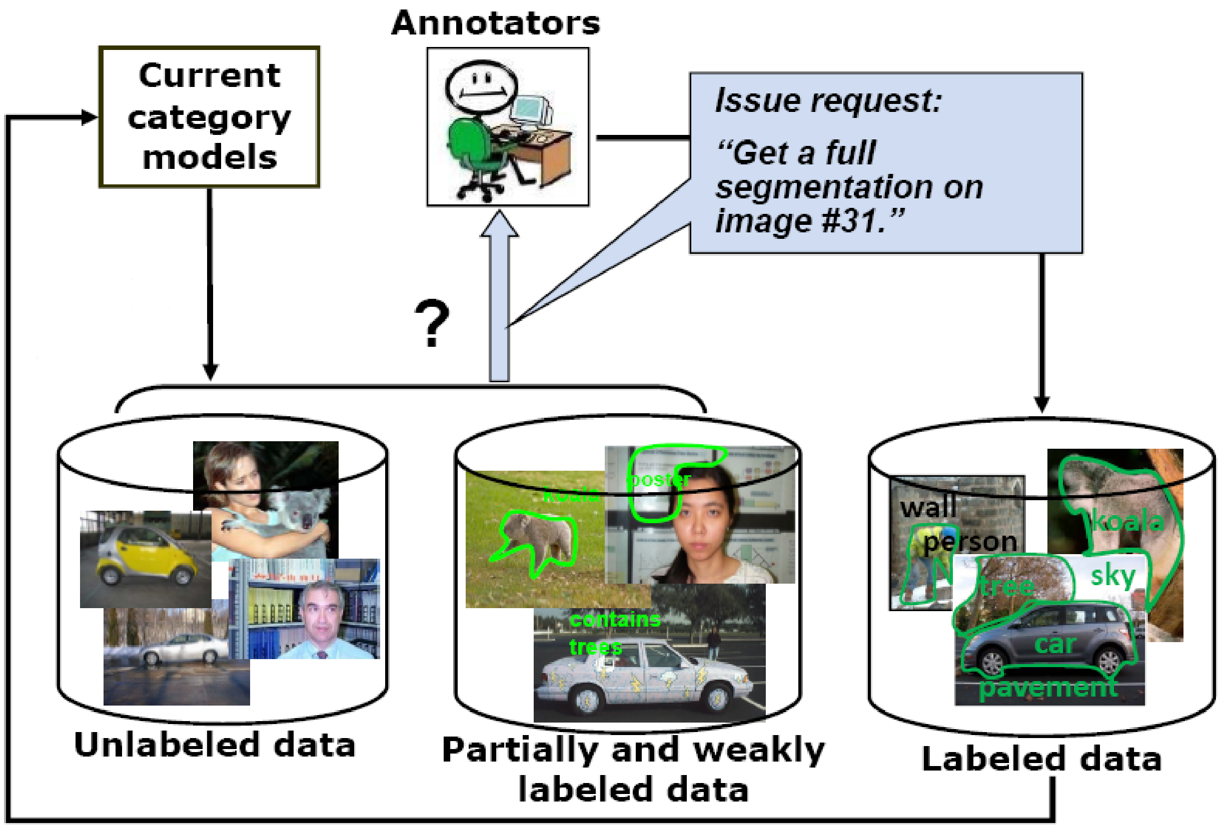
Other types of learning

- **Semi-supervised learning:** lots of data is available, but only small portion is labeled (e.g. since labeling is expensive)
  - Why is learning from labeled and unlabeled data better than learning from labeled data alone?



Other types of learning

- **Active learning:** the learning algorithm can choose its own training examples, or ask a “teacher” for an answer on selected inputs



Master Internships

- Internships are available in the THOTH group
- For research directions see

  http://thoth.inrialpes.fr

- If you are interested send an email directly to team members that you are interested to work with