Neural Networks and Deep Learning

Lecture 1: introduction

Jan Chorowski
Instytut Informatyki
Wydział Matematyki i Informatyki Uniwersytet Wrocławski

2017-2018

Slides taken from presentations by: A. Ng, P. Vincent, Y. LeCun, W. Zaremba, R. Salakhutdinov, L. Fei-Fei, A. Karpathy

Practical Information

- Website: http://ii.uni.wroc.pl/~jch/teaching/fall2018-2019/neural-networks-fall-18
- Free on-line Textbooks:
  - http://www.deeplearningbook.org/ -> You can also buy a printed version!
  - http://neuralnetworksanddeeplearning.com/
- Good lecture notes for 50% of material in this course: http://cs229.stanford.edu/
- More advanced, very good courses on deep learning:
  - https://ift6266h15.wordpress.com/
  - http://cs231n.stanford.edu/
  - http://cs224d.stanford.edu/
- Good paper textbooks:
  - Goodfellow, Bengio, Courville „Deep Learning" http://www.deeplearningbook.org/
  - Murphy, Machine Learning: a Probabilistic Perspective http://www.cs.ubc.ca/~murphyk/MLbook/
  - Bishop, Pattern Recognition and Machine Learning (PRML) https://www.microsoft.com/en-us/research/people/cmbishop

Why are you here?

What do you want to learn?

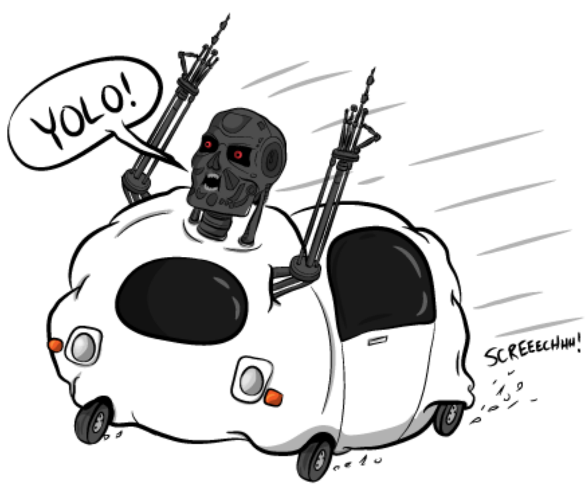
1. Machine Learning: using data to teach machines useful skills

2. Neural Networks: good Machine Learning models

Where is machine learning?

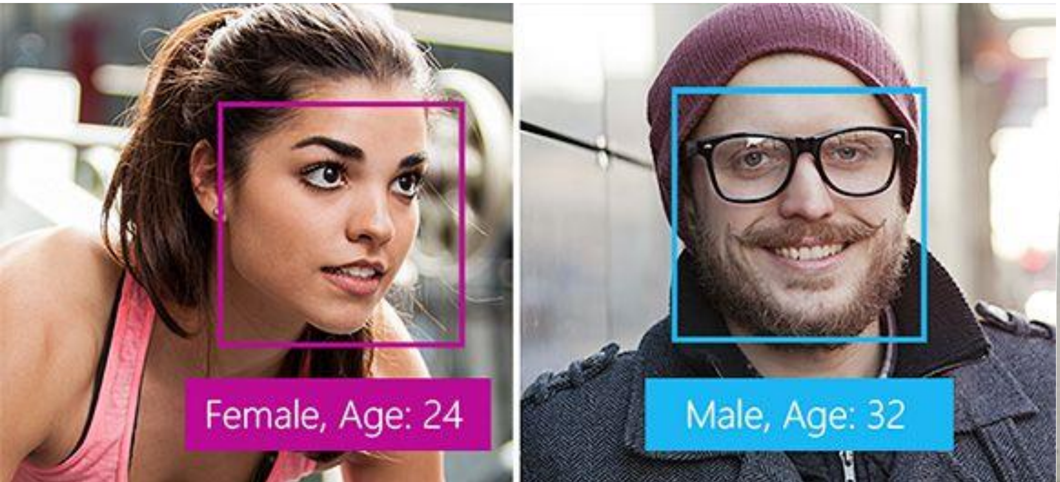
Everywhere!

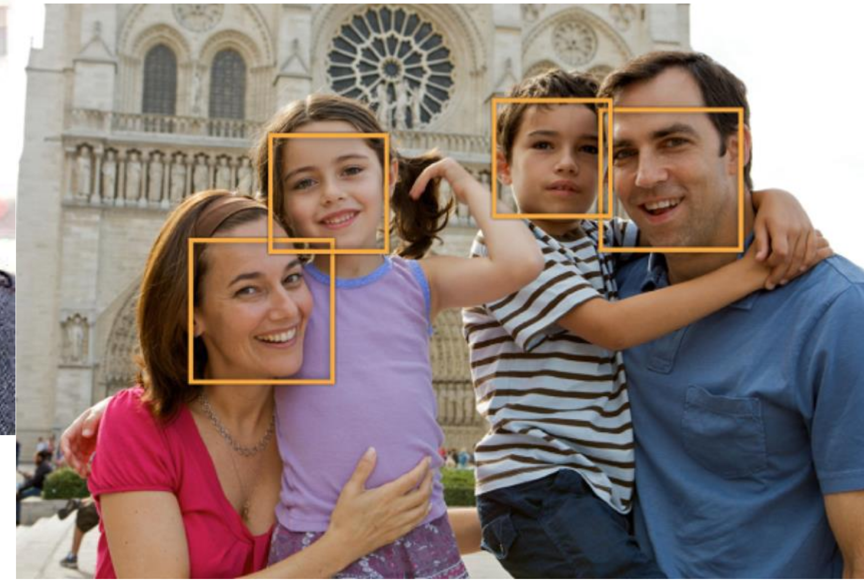
- web search ☺
- and ads ☹
- recommendations
- Self-driving cars



http://theoatmeal.com/blog/google_self_driving_car

ML example: face detection

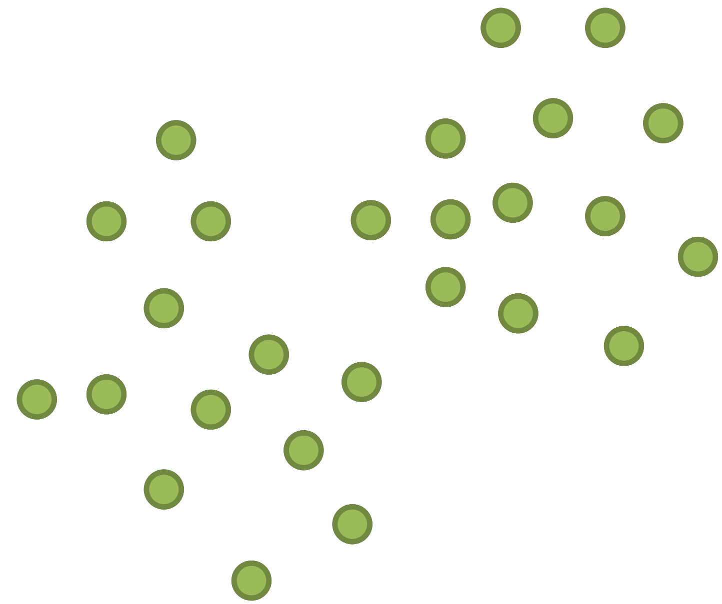




Source: Microsoft and Apple face detection API documentations, wykop.pl

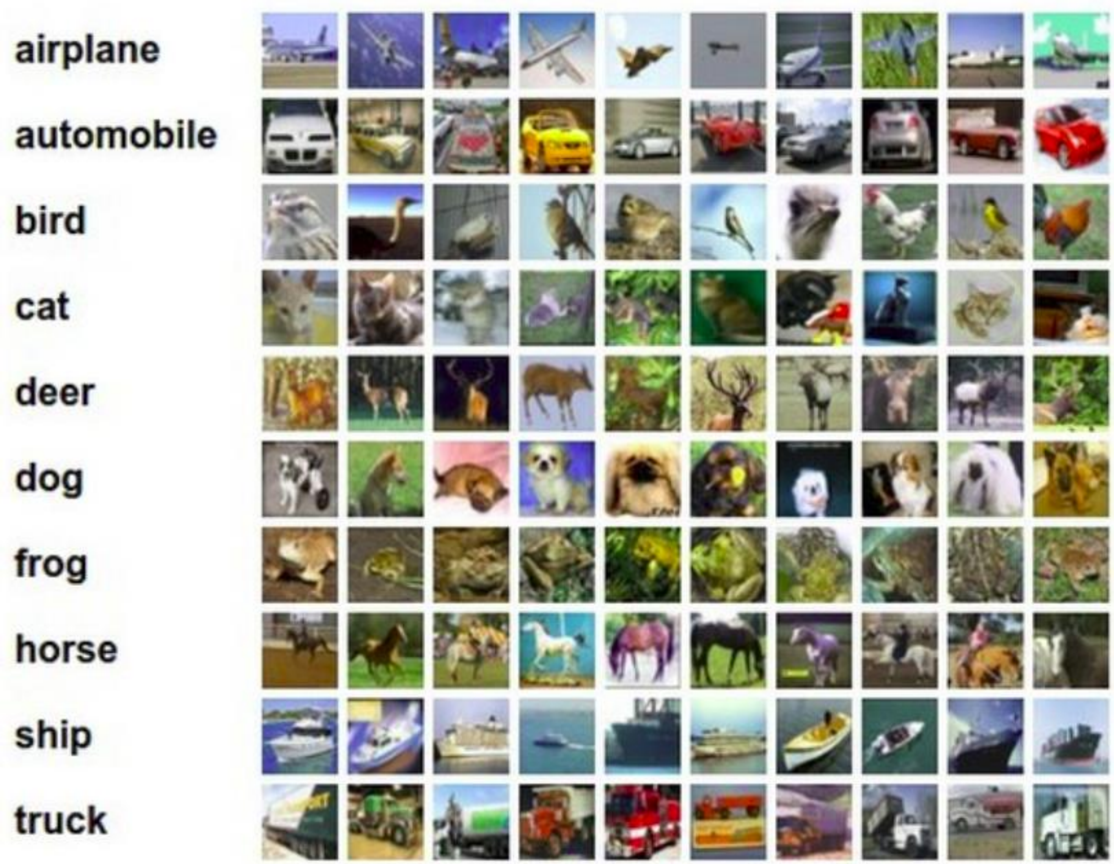
Types of learning

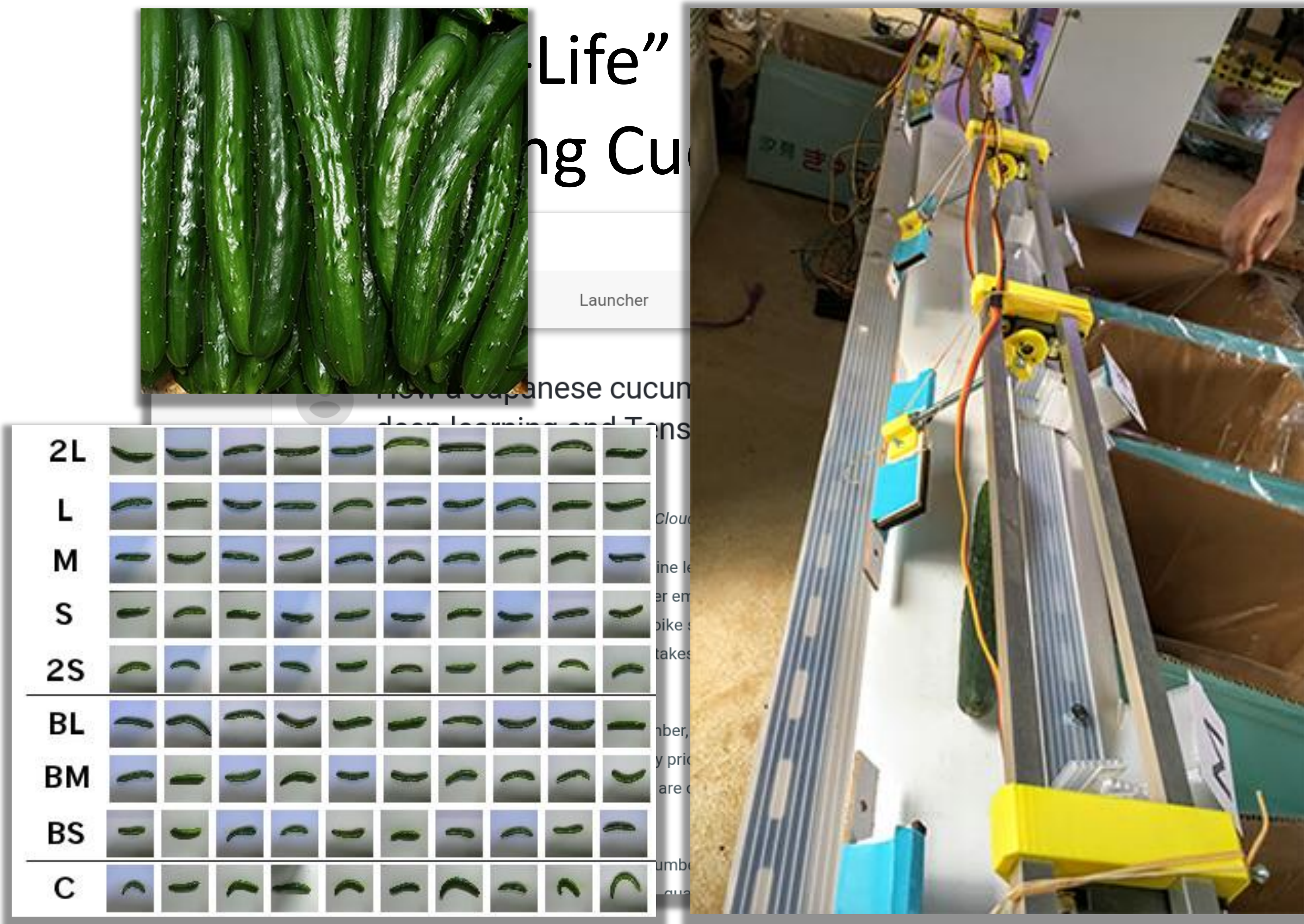
- Supervised:
  - Learn an input-output relation, instant feedback
- Unsupervised:
  - Learn the density of data
  - Learn how to generate data
  - Group data into clusters
- Reinforcement:
  - Reward information given after a series of actions
  - Think of learning strategy in games



Supervised learning examples

- Assign small images to one of ten categories





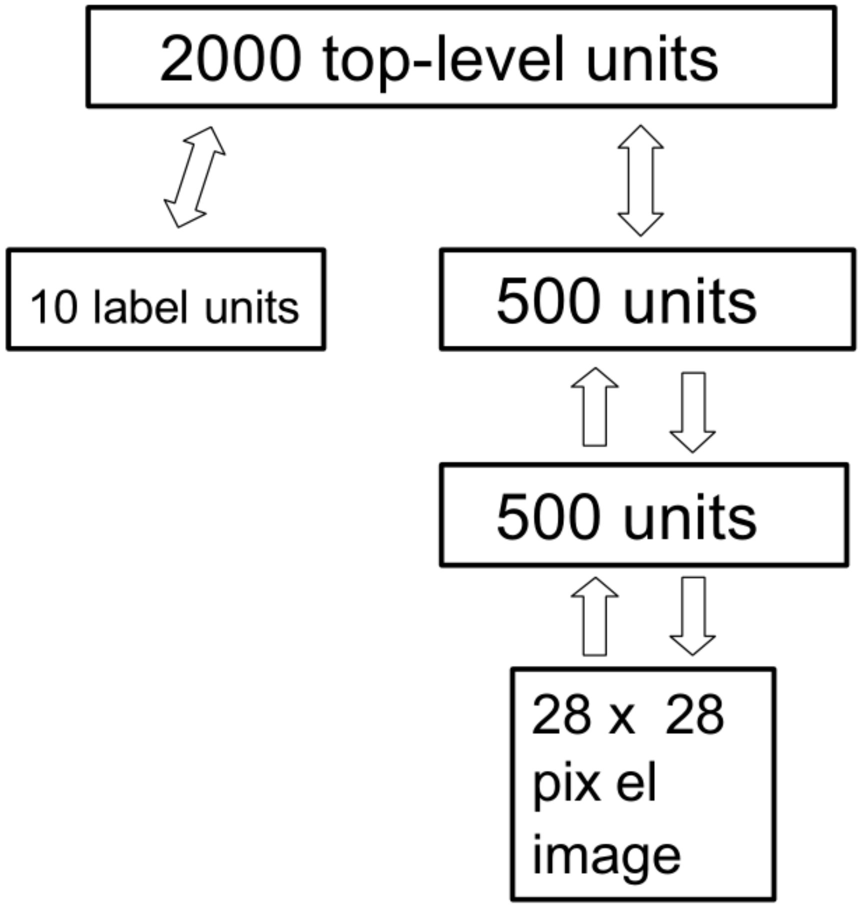
Source: https://cloud.google.com/blog/big-data/2016/08/how-a-japanese-cucumber-farmer-is-using-deep-learning-and-tensorflow

Unsupervised learning example

- Generating album covers

Image by Alec Radford

Digit Generation



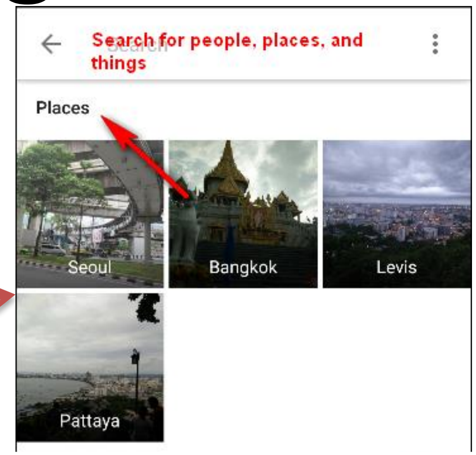
Demo: http://www.cs.toronto.edu/~hinton/adi/index.htm

ML Example: Google Photos app

Algos and Data

Sup. learning

Unsup. learning



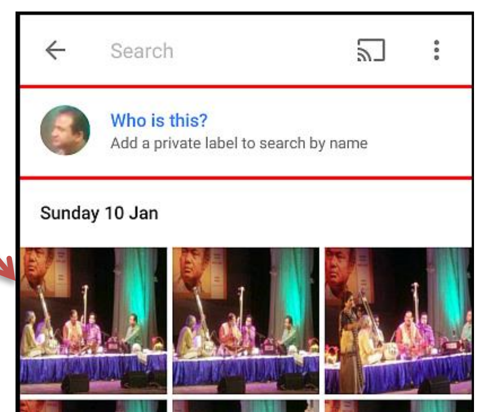
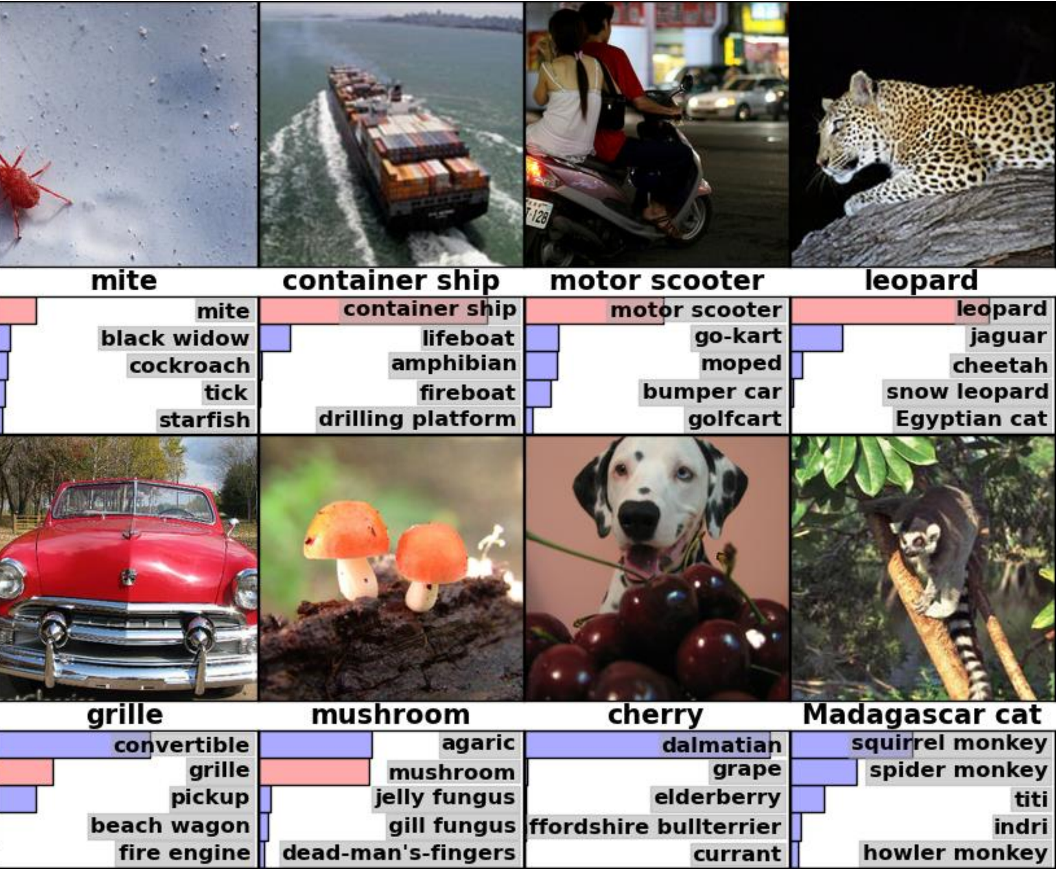
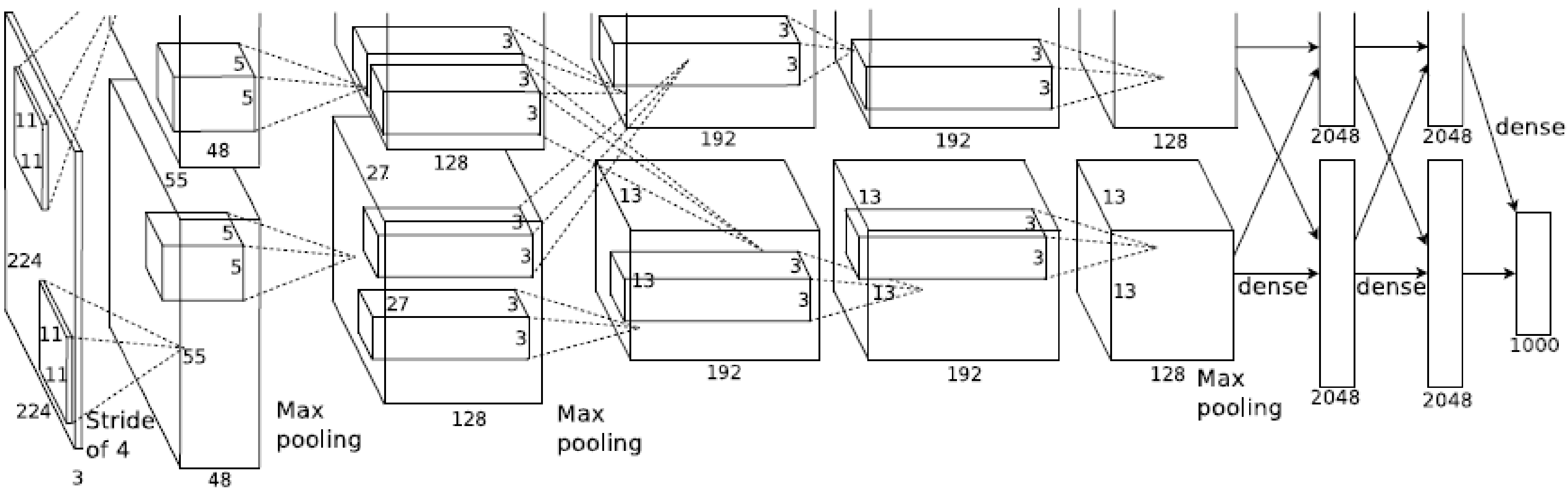


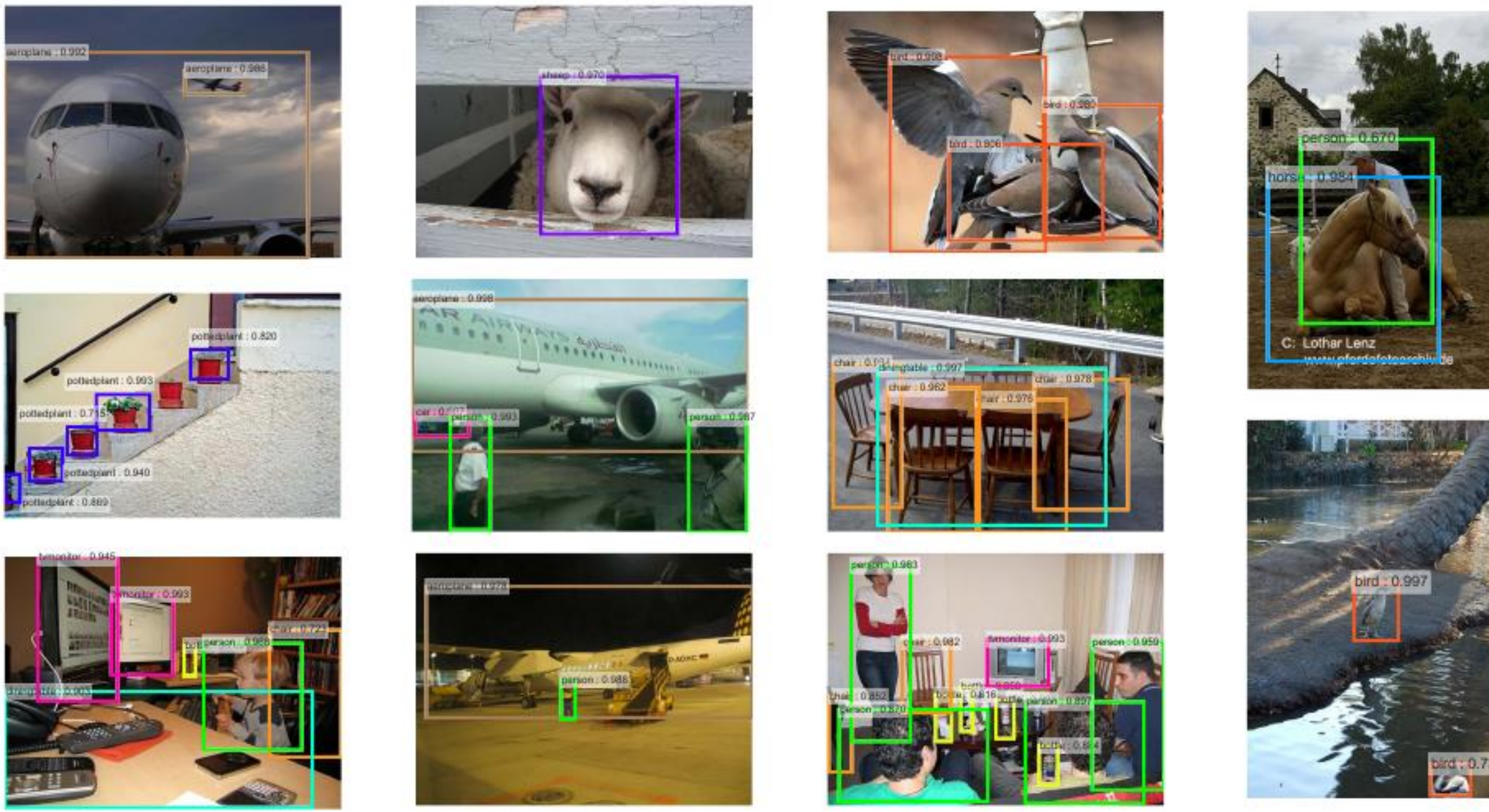
Image recognition





Object Detection

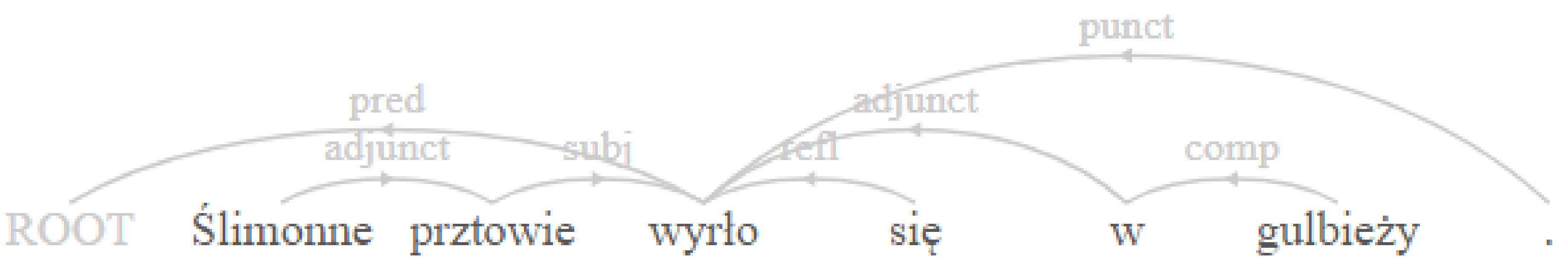
https://www.youtube.com/watch?v=WZmSMkK9VuA



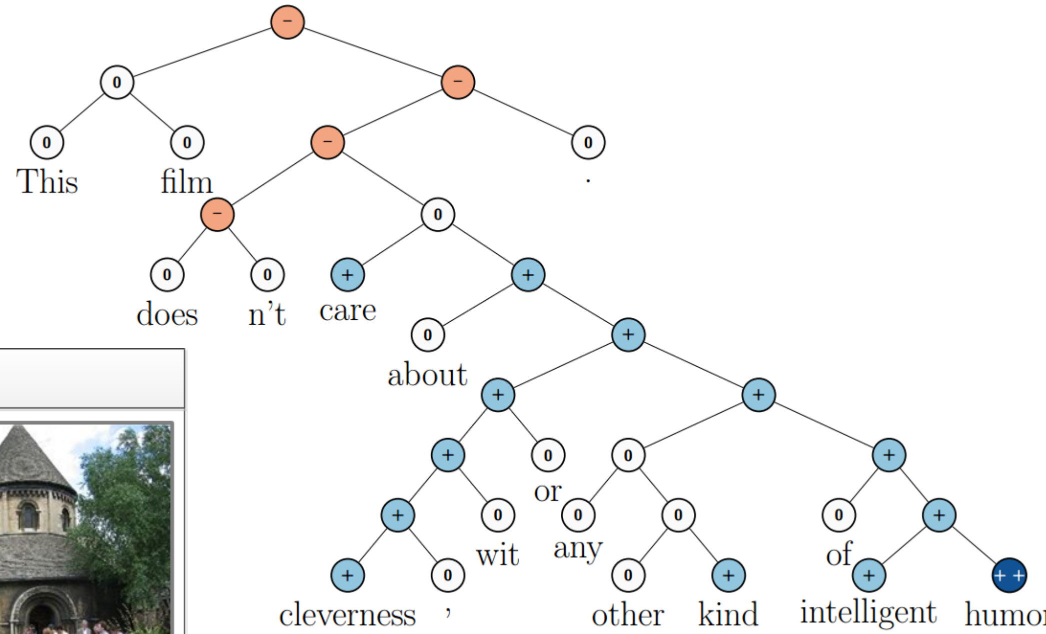
(Ren, He, Girshick and Sun, „Faster R-CNN", 2015)

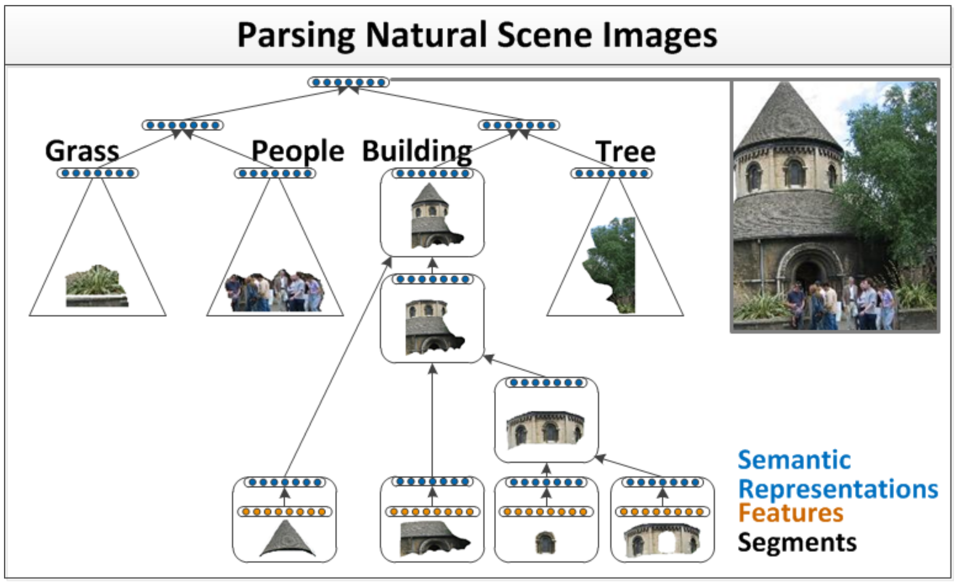
Natural Language Processing

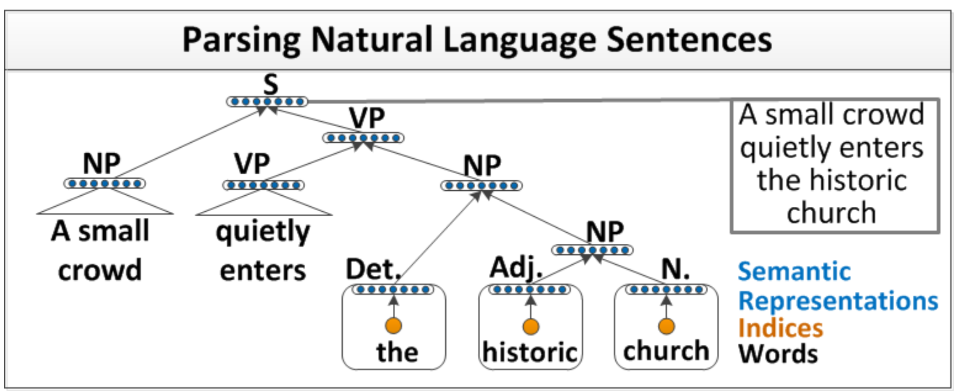
http://zapotoczny.pl/parser/



NLP, embeddings



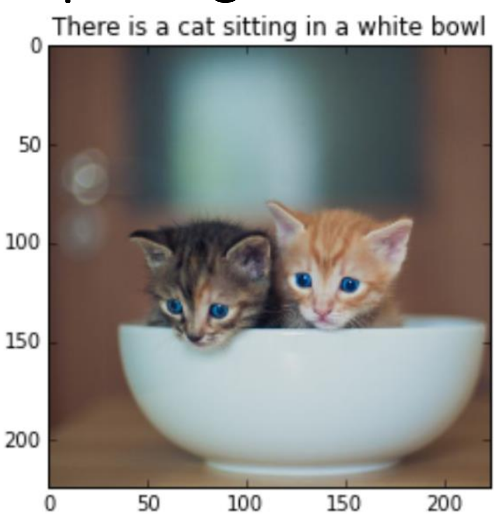


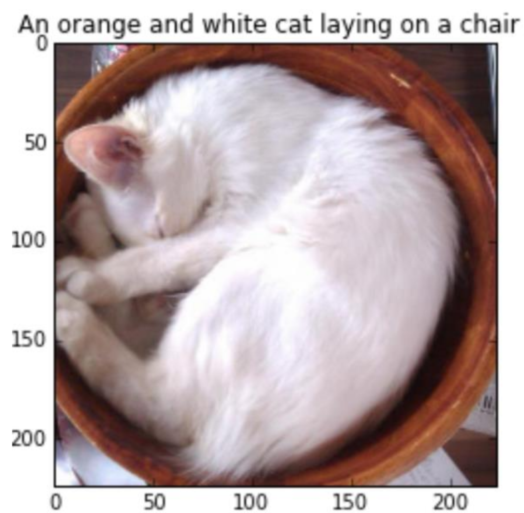


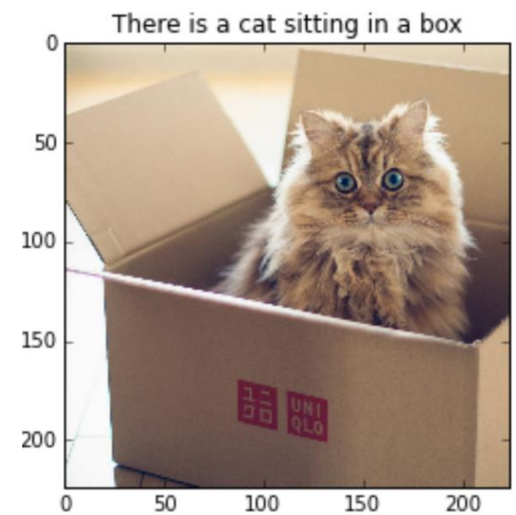
Using both language and images

Putting images and tags into the same vector space
(Kiros, Salakhutdinov, Zemel, TACL 2015)

Caption generation (Xu et al ICML 2015)





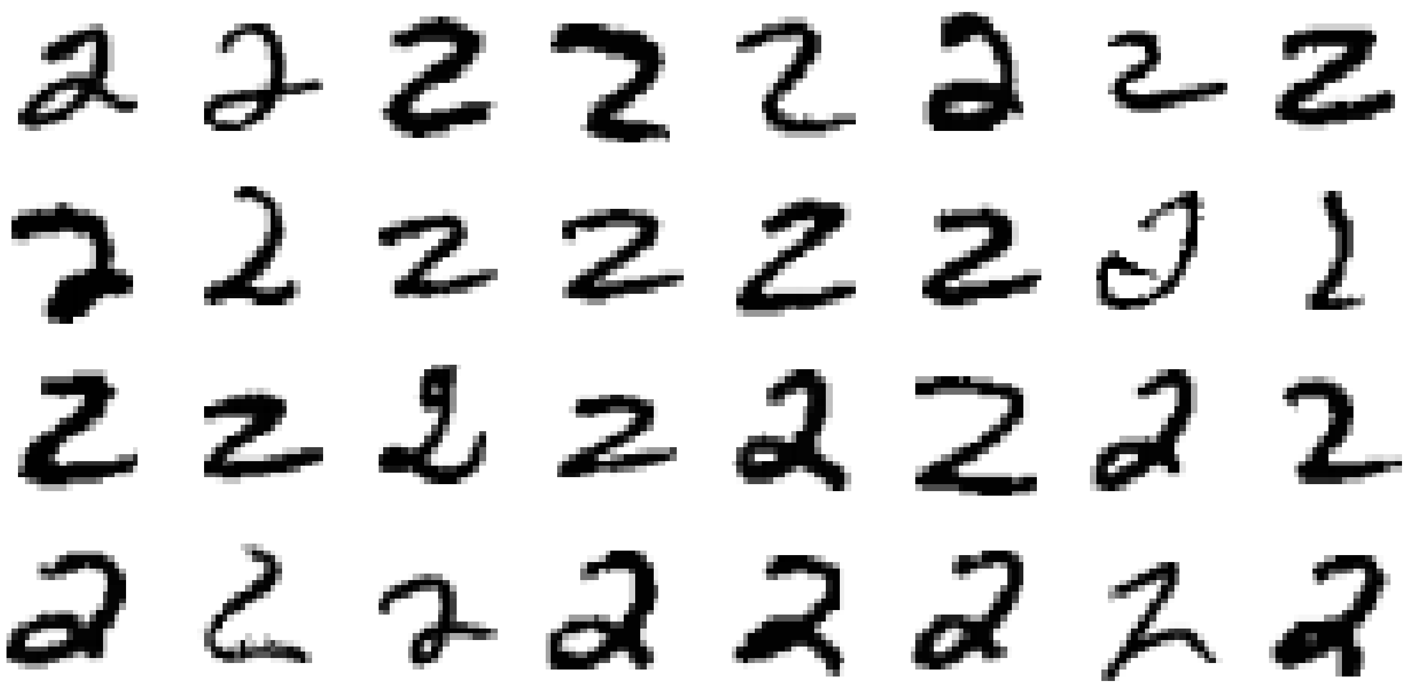


How are these problems solved?

1. Take **data** and a **learning algorithm**
2. The algorithm discovers **patterns** in the data and produces a **model**
3. Query model to ask questions about the data
   - „Is there any face in the image?”
   - “Is this review favorable?”
   - „What is the object in the image?”
   - “What caption is most likely given this image?”

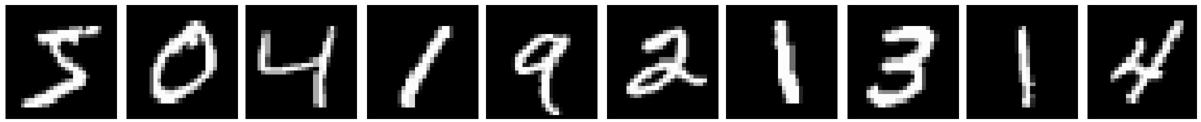
Example: Digit Recognition

- Task: recognize handwritten digits
- Input: images 28 x 28 pixel values ($[0,1]^{784}$)
- Output: {0,1,…,9}

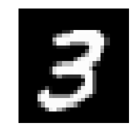


The easiest algorithm

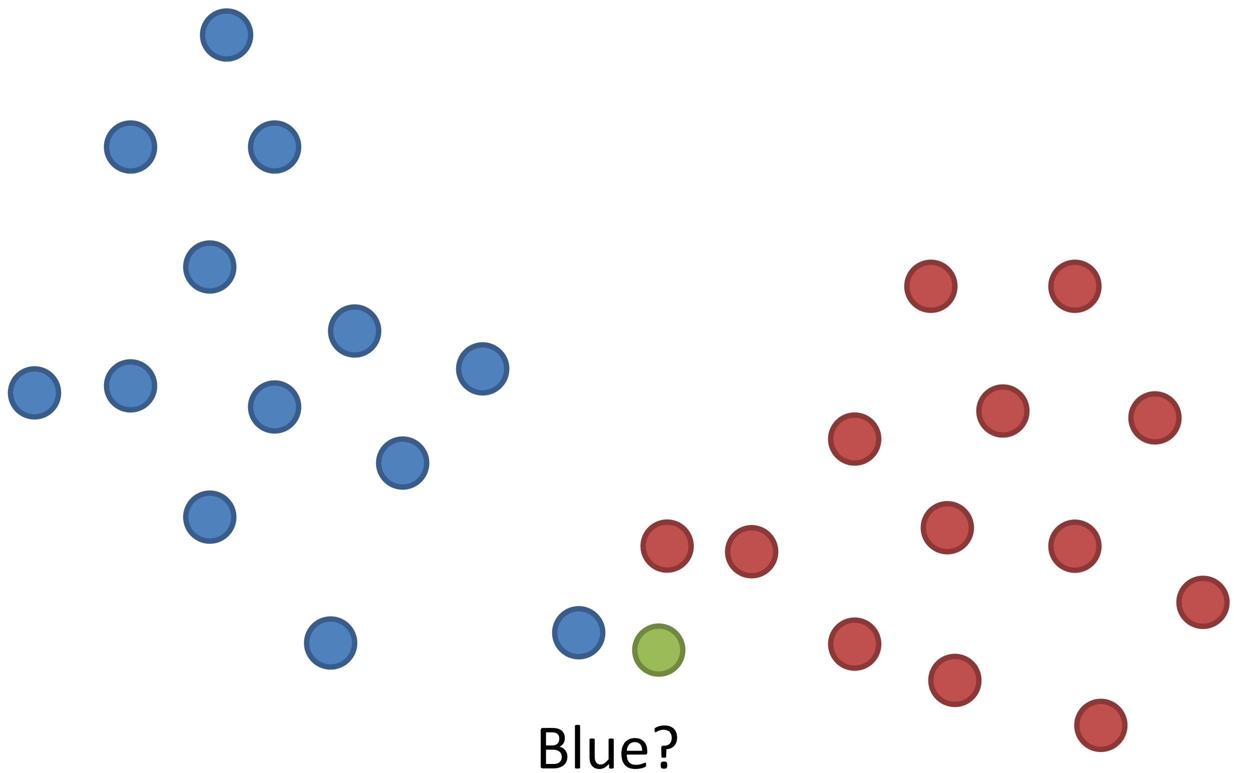
- Collect examples – pairs of (input, output):



5 0 4 1 9 2 1 3 1 4

- To classify a new instance  :
  – Find the most similar element of the train set 
  – Return its label
- Seems too easy
  – It works well when we have enough data

KNN (k Nearest Neighbors)



Design decision: choose “k” – how many neighbors to use?

Few – sensitive to outliers, Many- very smooth classification boundary

KNN uses: recommendations

- Items are similar if users rate them similarly
- Training (offline):
  - Represent each item by its scores
  - Compute the distance between scores
- Recommendation generation (online):
  - Find nearest neighbors for each item recently browsed/put into the cart

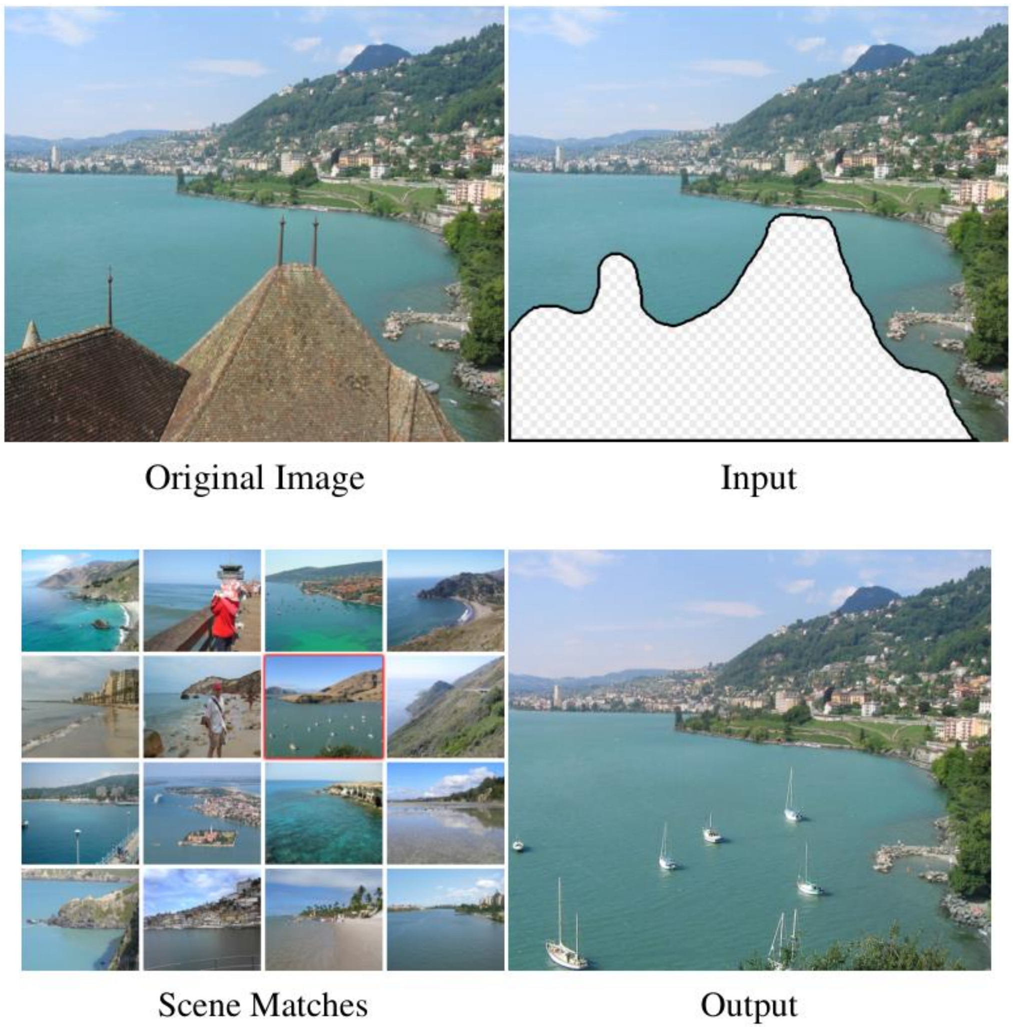
Amazon patented this:

US007113917B2

(12) **United States Patent**
Jacobi et al.

(10) **Patent No.:** **US 7,113,917 B2**
(45) **Date of Patent:** ***Sep. 26, 2006**

Nearest Neighbors use: inpainting



1. Find similar photos
   – Use Gist descriptors
2. Match and copy

Works poorly with 1000 images…

But is good with >1000000

Hays, Efros „Scene Completion Using Millions of Photographs”

Alternative realities

Input

Alternative inpainting results

kNN for dialogs

A system Adrian Łańcucki coded based on Reddit.
Real (cherry-picked) conversation from Festiwal nauki:

User: hi

Bot: Hi.

User: i like you

Bot: Your in if u want.

User: lmao

Bot: Ayy lmao.

User: ayyy

Bot: Who are you?

User: im a nurse

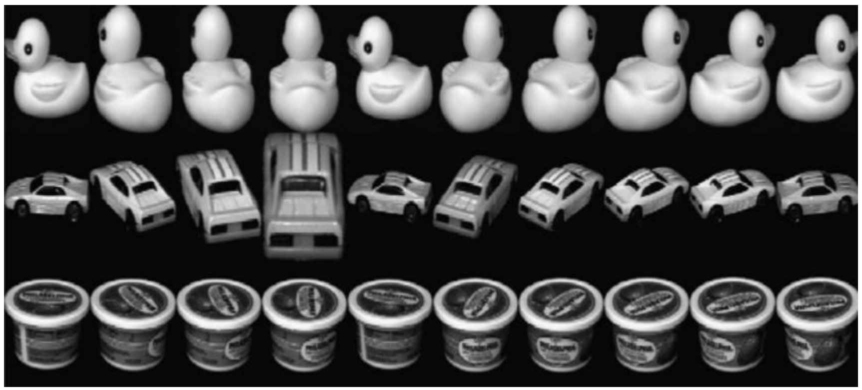
Bot: Gott im Himmel. You have become the host.

User: and what are you ?

Bot: I am just a talking robot. But I like my job.

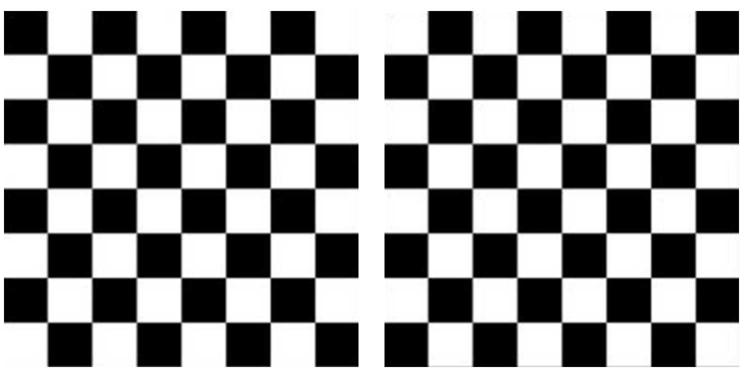
User: thats great

KNN – what makes images similar?



Quite similar

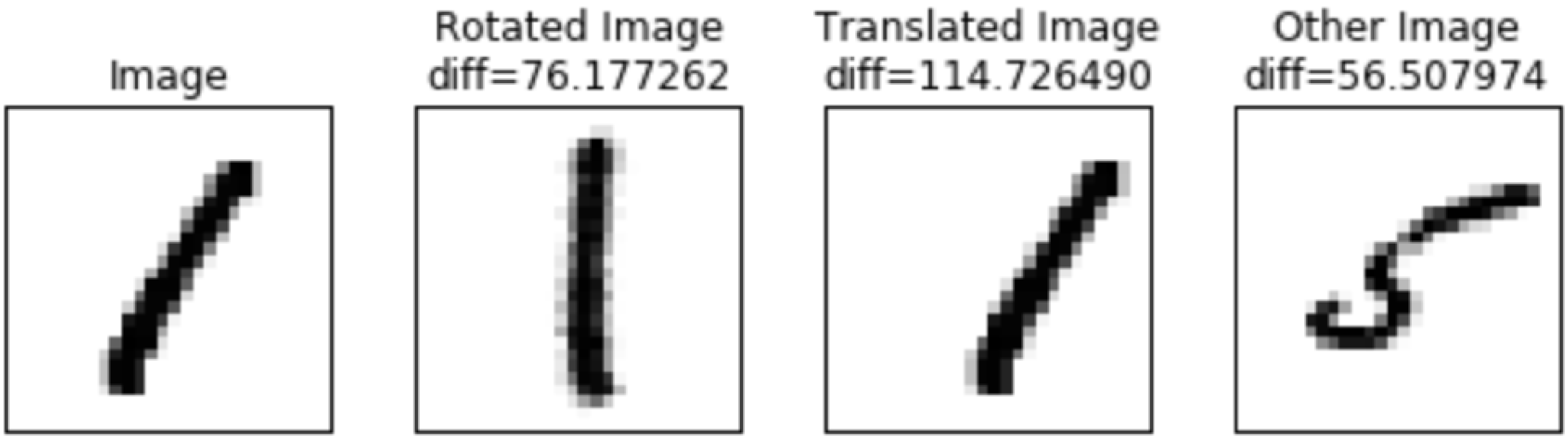
Not quite similar



Opposite

KNN – what makes images similar?

Pixel-wise difference can be large for visually similar images



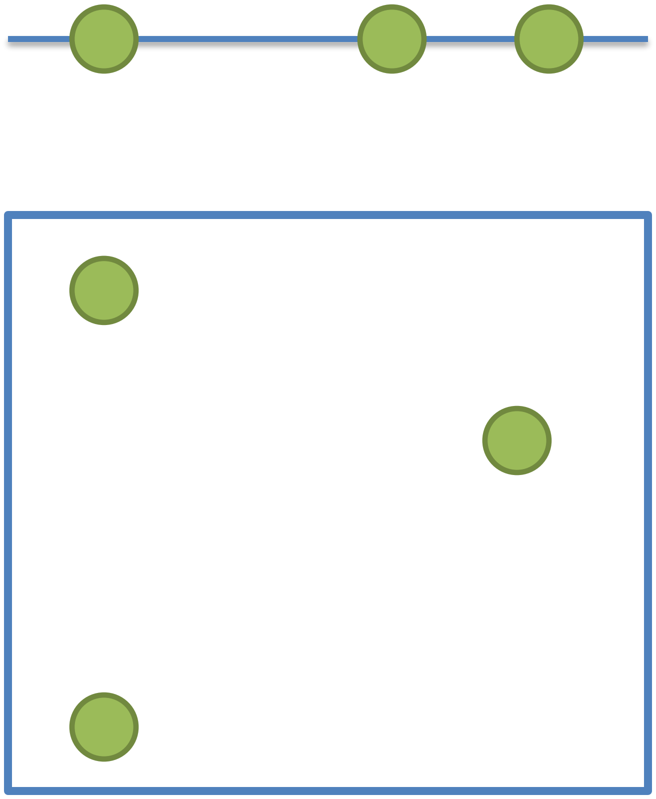
KNN requires lots of data

- Consider recognizing postal codes (5 digits)
- Will this work if we treat each postal code as a single image (10^5 combinations)?
- If we treat each digit separately, how to accommodate for:
  - Uniform style (slant, thickness) of digits in one code.
  - Segmentation, joins, overlaps of digits.

Curse of dimensionality

In highly dimensional spaces, things are dissimilar (far away from each other):

- Let N be the number of points.
- The average distance between points grows with data dimensionality.
- In other words, there are fewer neighbors within a radius from each point.



Humans see differently

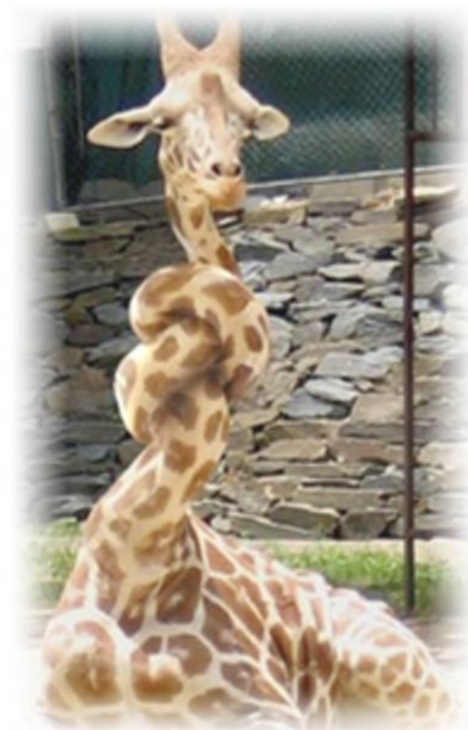
Change blindness

- Humans are blind even to very drastic changes in images.
- We clearly treat things as the same even if the pixel-wise differences are huge!
- Example:

  https://www.youtube.com/watch?v=1nL5ulsWMYc

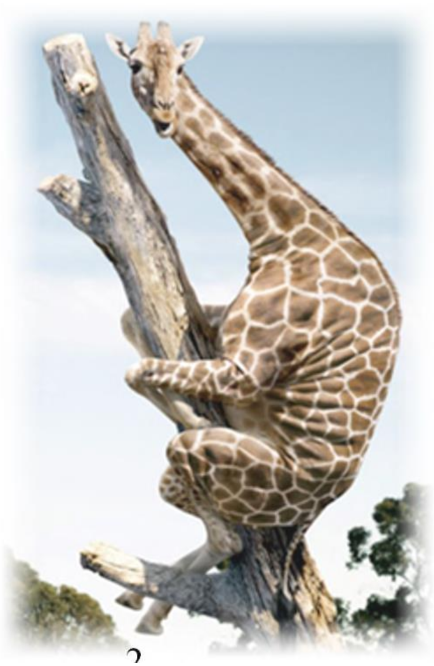
  https://www.youtube.com/watch?v=FWSxSQsspiQ

Learning is about invariants and generalizations



Thomas Serre 2012, slide by Nando de Freitas

We want a hierarchical model



Low-level concepts:

- Edges
- Points
- Gradients
- ....

Higher-level:

- Textures:
  - Spots
  - Stripes
- Geom. shapes:
  - Corners
  - Circles
  - Lines

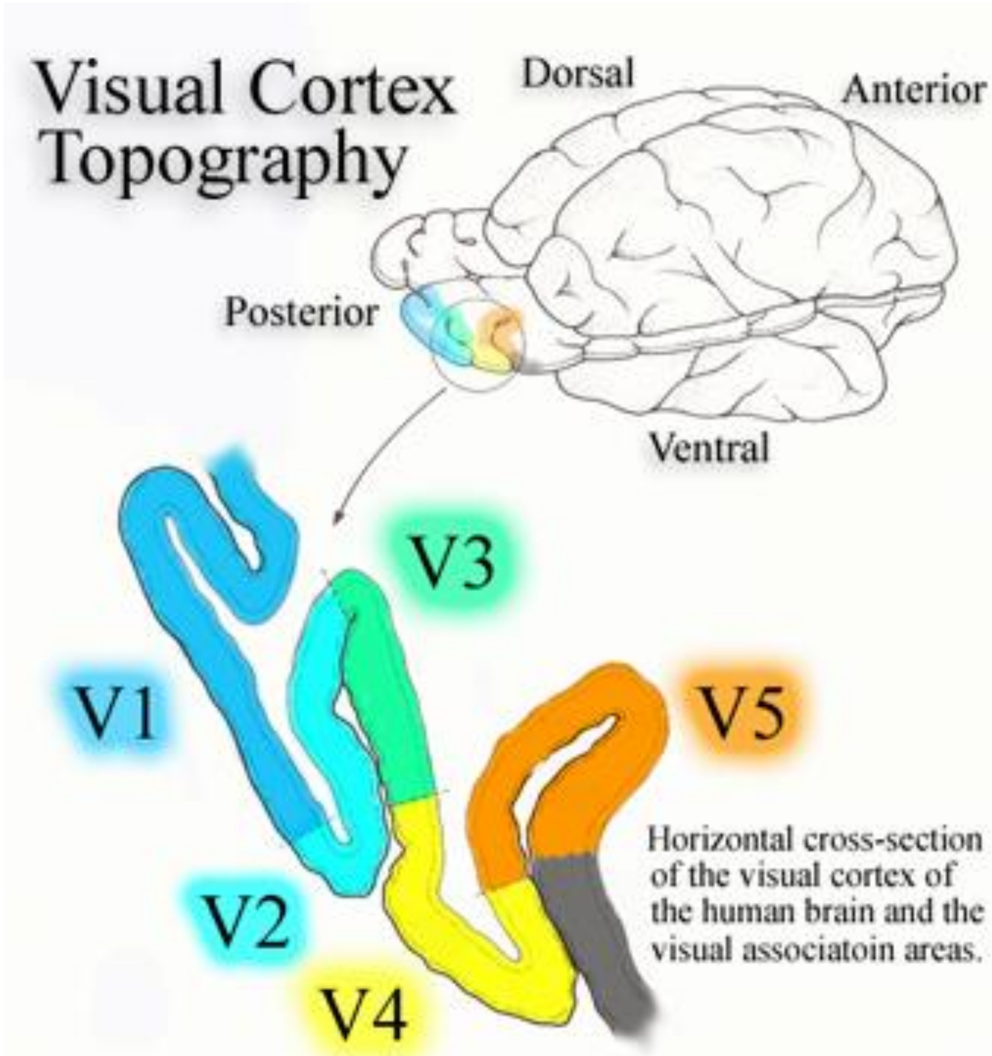
Object parts:

- Paws
- Legs
- Necks
- Heads
- …

High-level Objects:

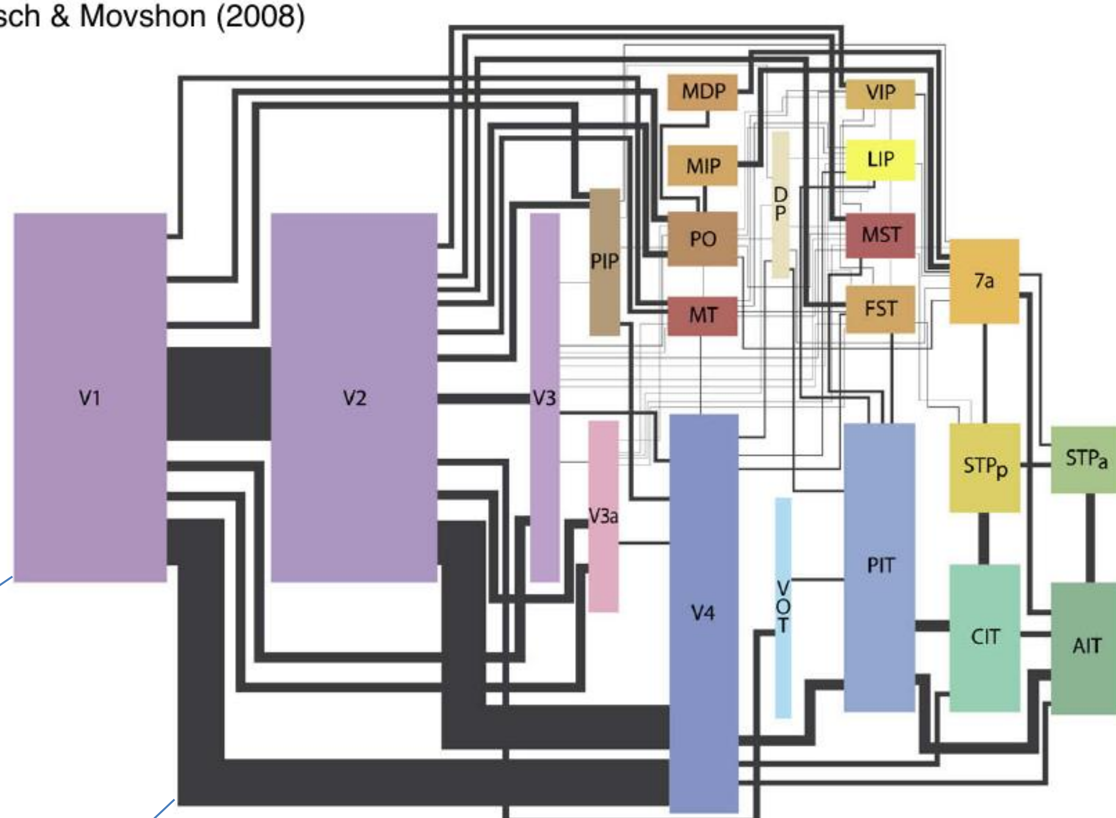
- Lions
- Giraffes
- Zebras
- …

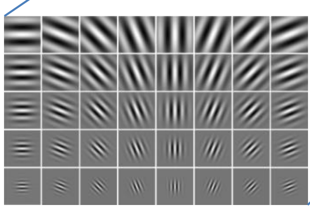
Brain & Visual Cortex



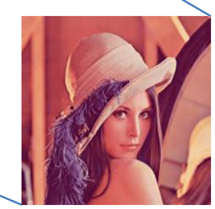
Visual Cortex Diagram

Wallisch & Movshon (2008)

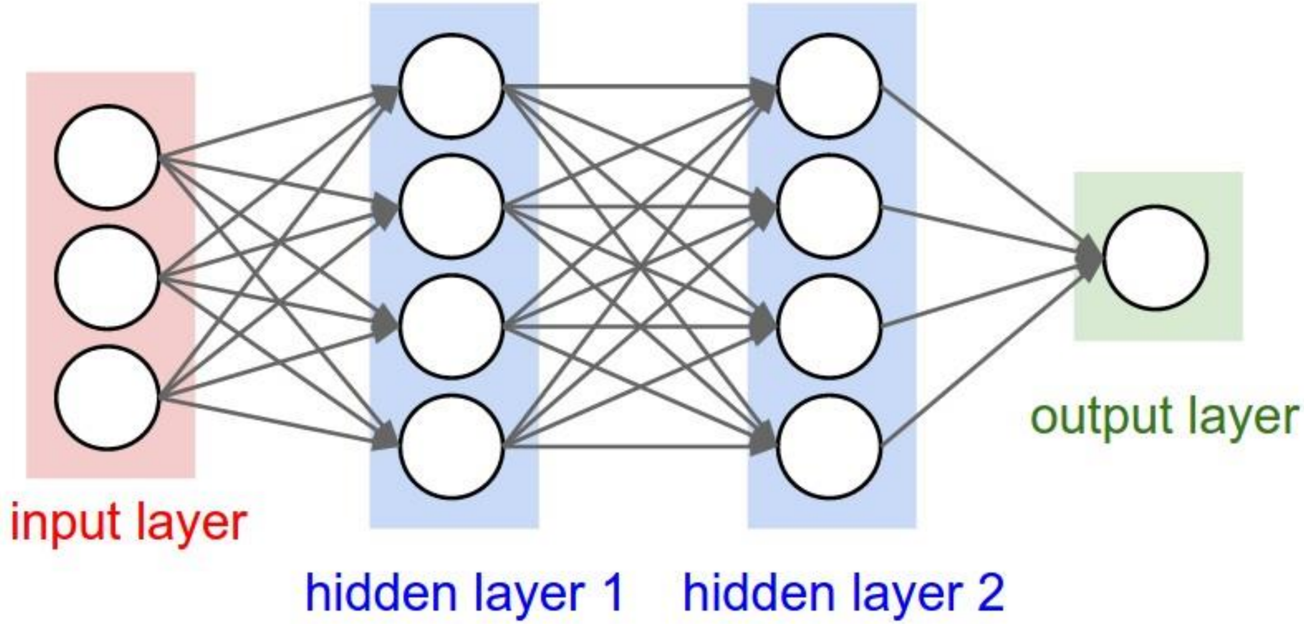




'Gabor filters' · · ? · · objects · faces



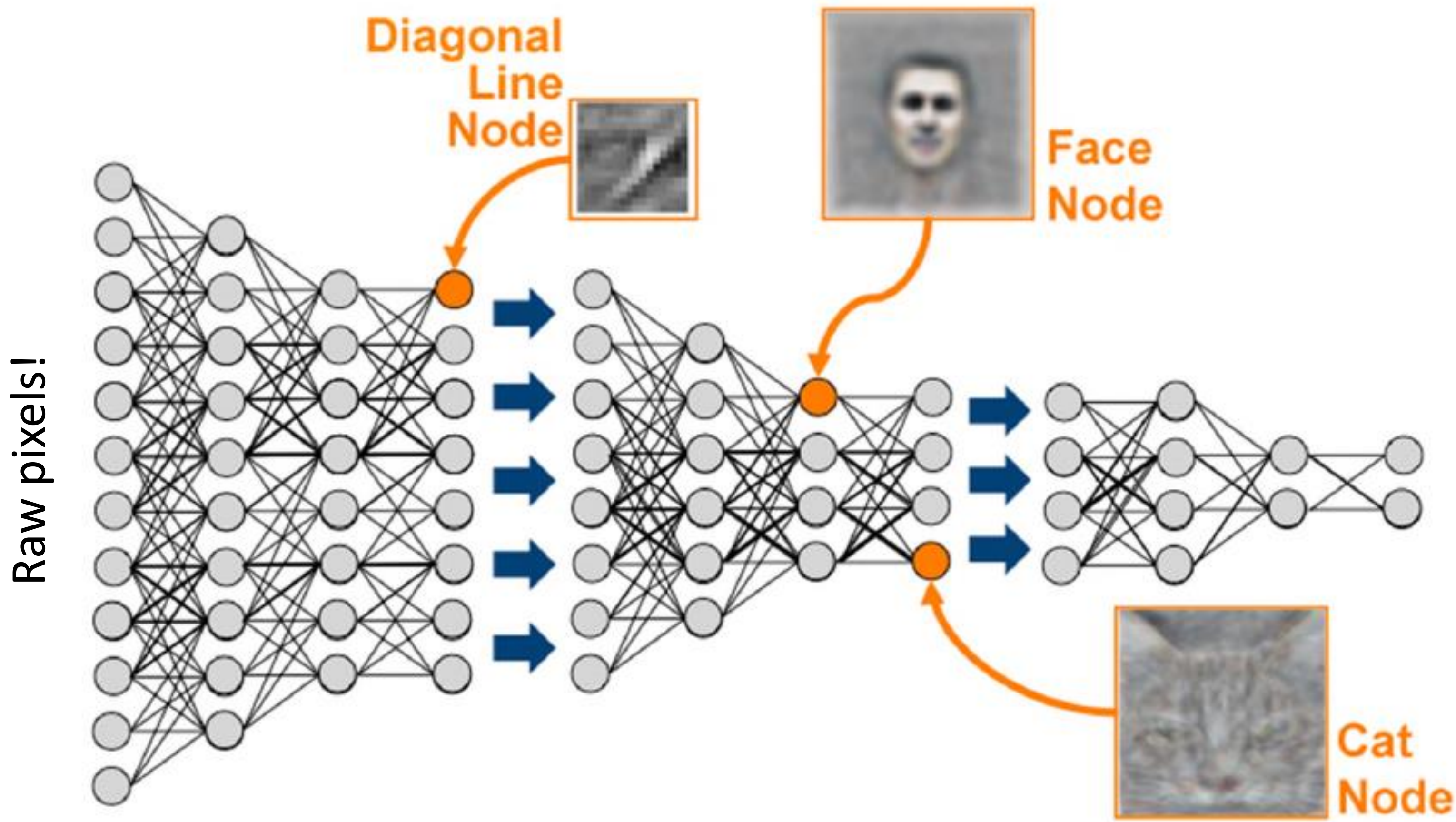
Neural networks



- A neuron detects some patterns in its inputs – combinations that cause it to fire
- When assembled into a network, neurons deep in the network react to patterns composed of more primitive parts

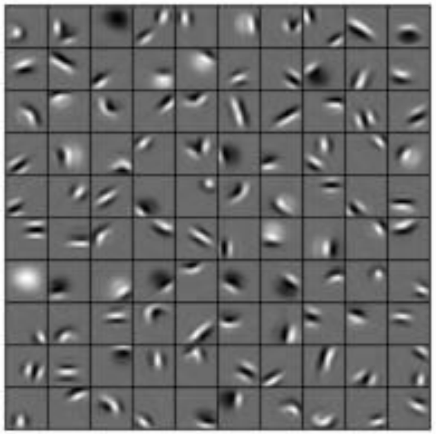
Neural nets learn hierarchies!

Google trained a network on YouTube videos. The net developed units detecting persons and cats!

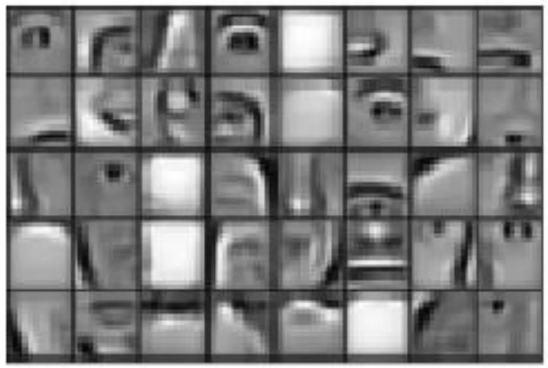


Q. Le et al. „Building high-level features using large scale unsupervised learning”

Neural nets learn hierarchies!



First layer



Second layer

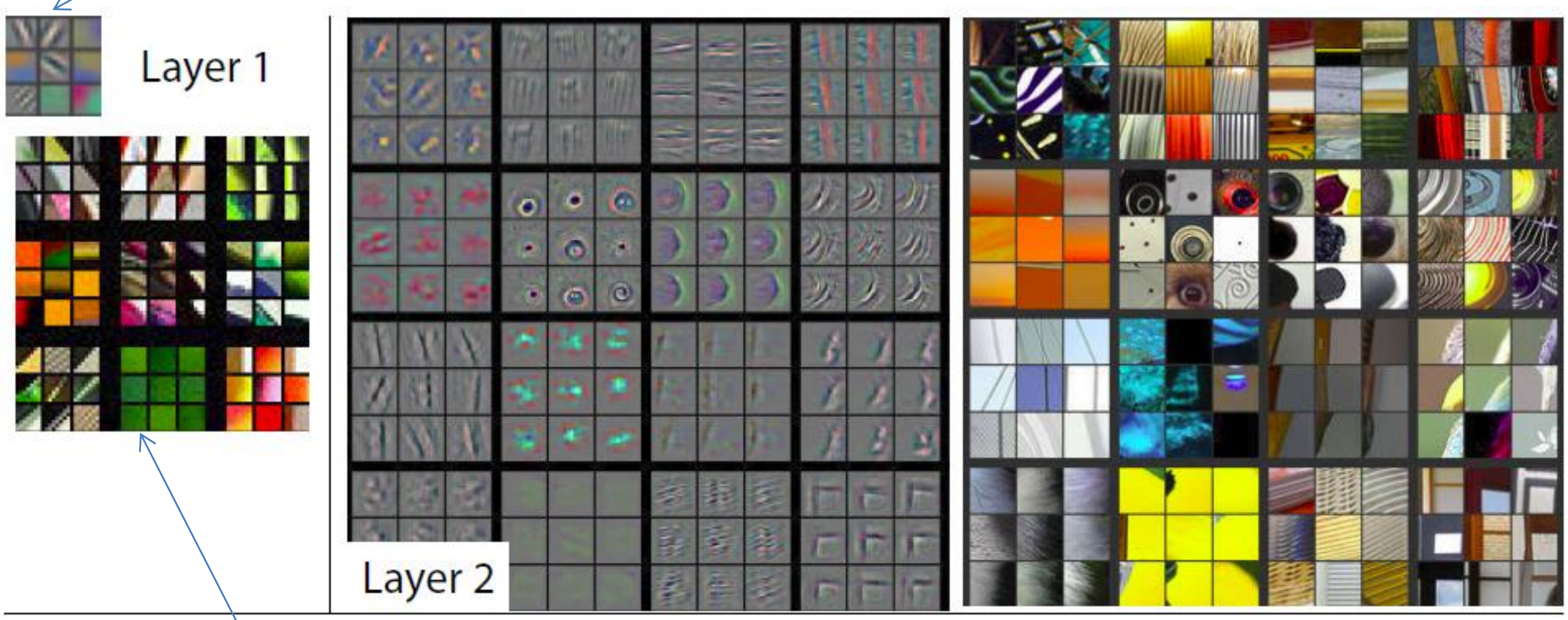
Third layer

Hierarchical features learned from a dataset of face images

(Lee et al., „Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks”)

Low-level features

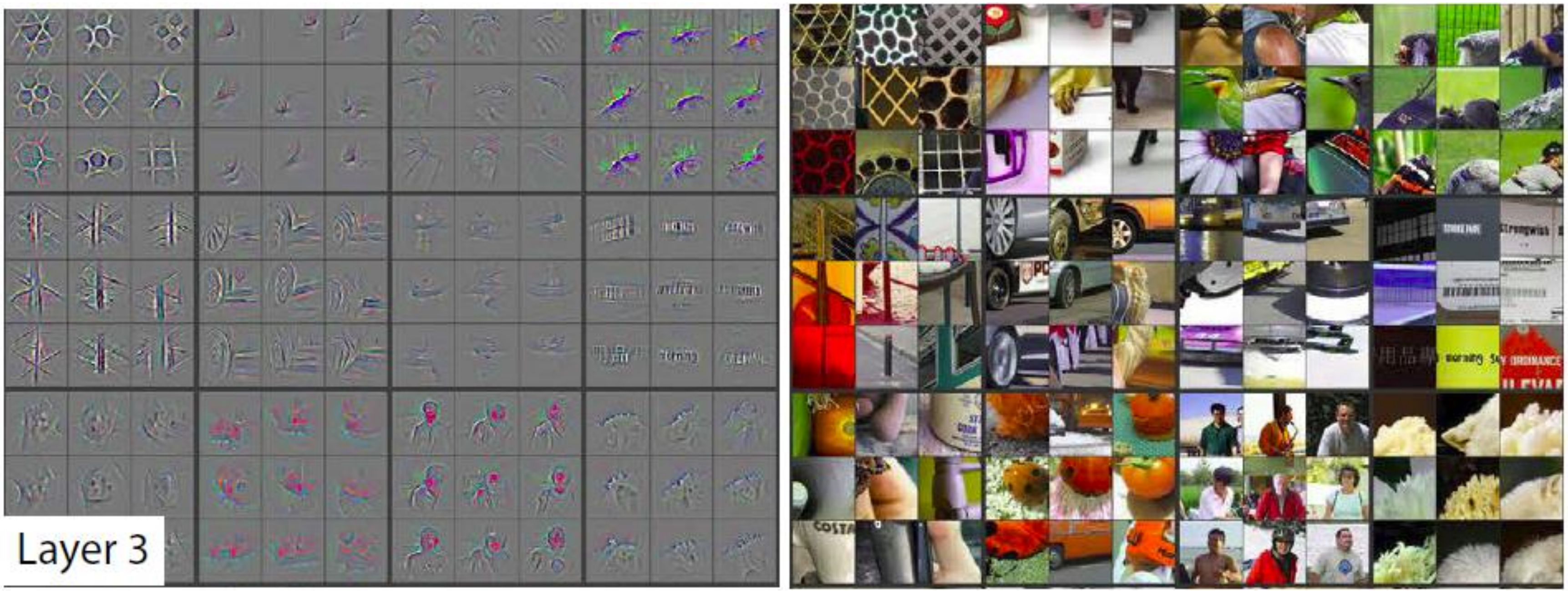
What the neuron (feature-detector looks for)



What images are selected by the neuron

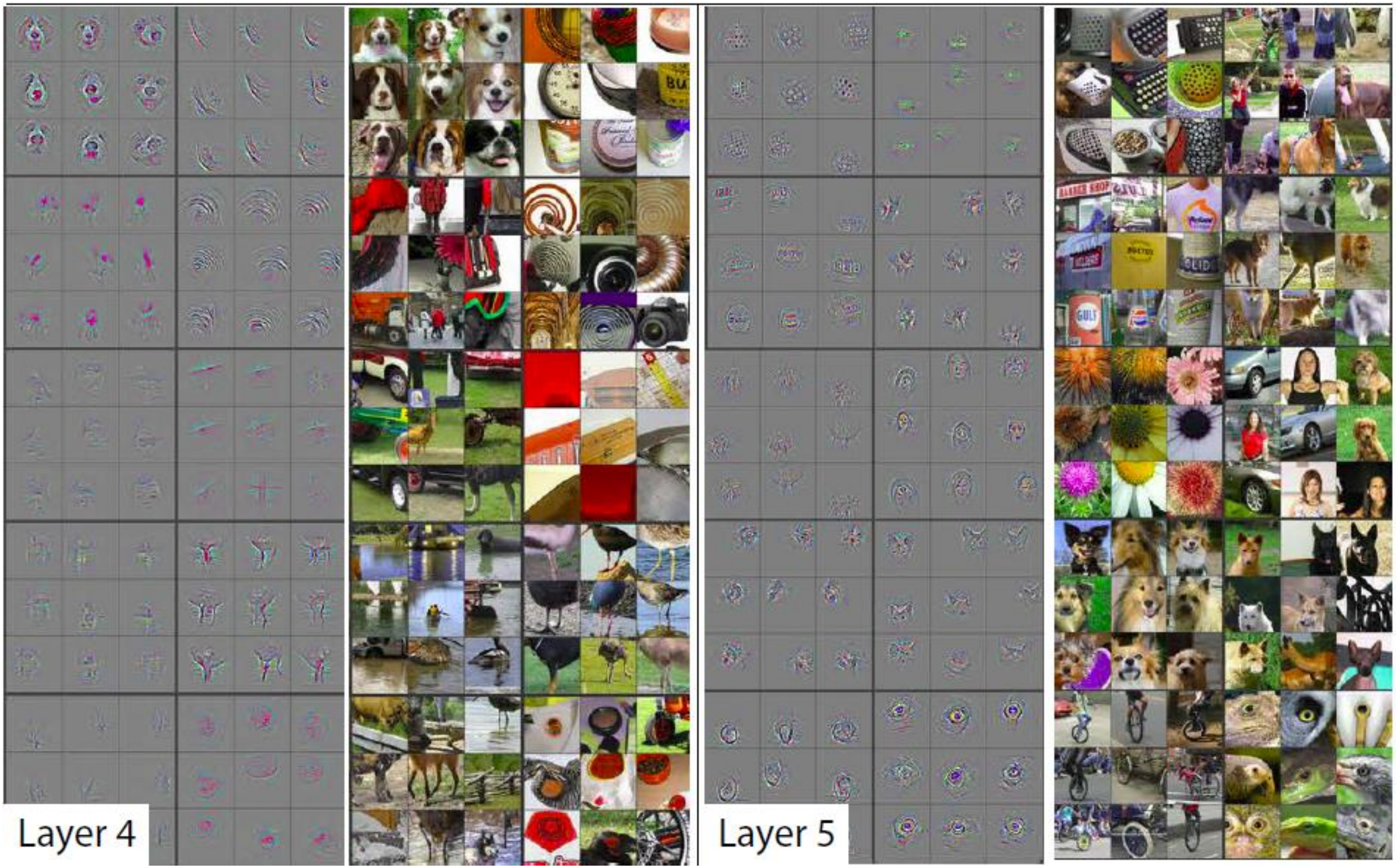
M. Zieler, “Visualizing and Understanding Convolutional Networks”

Mid-level features



M. Zieler, “Visualizing and Understanding Convolutional Networks”

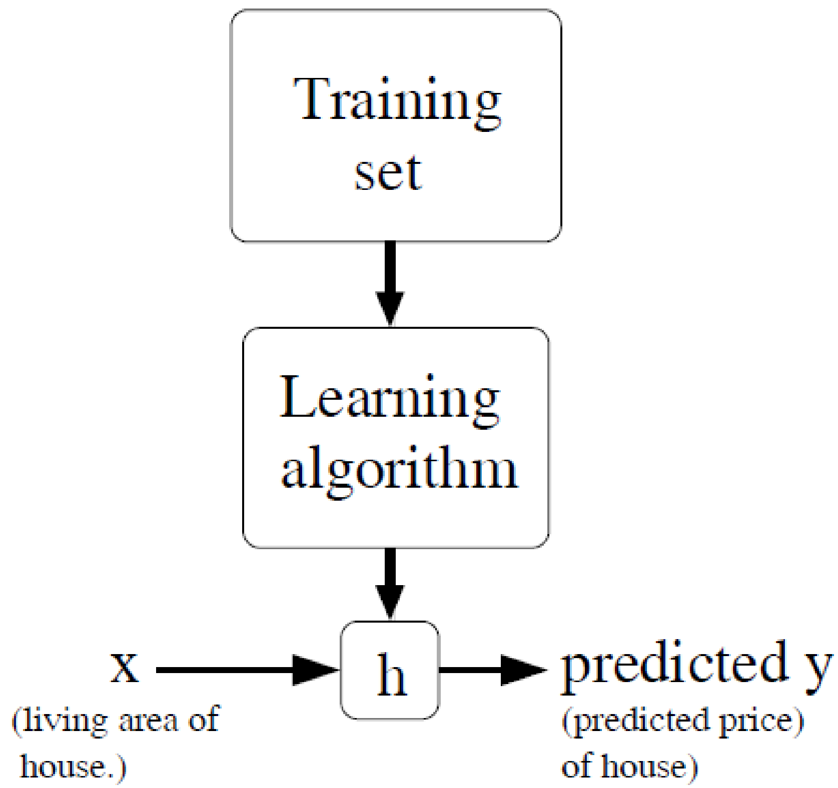
High-level features



M. Zieler, “Visualizing and Understanding Convolutional Networks”

Quick summary

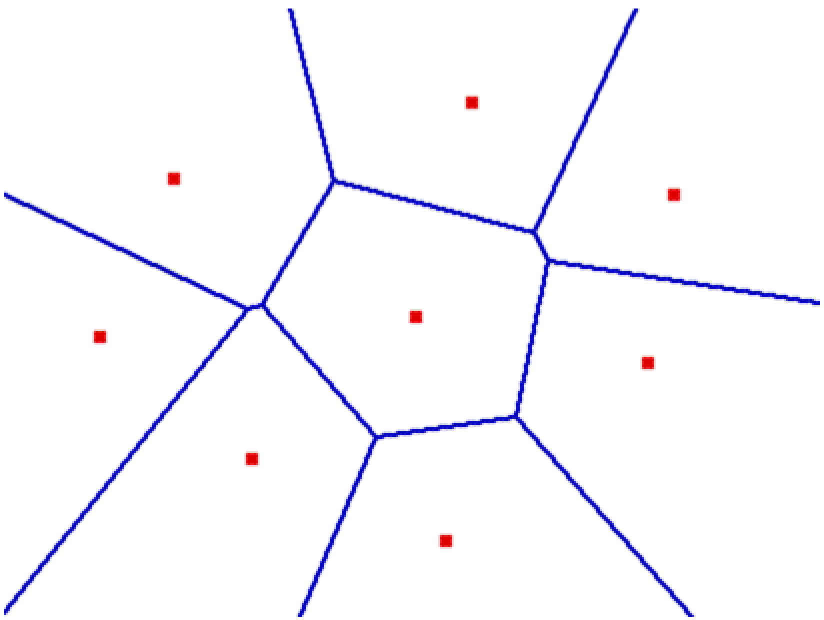
- ML algorithms distill data into models.
- We will call this learning from examples.



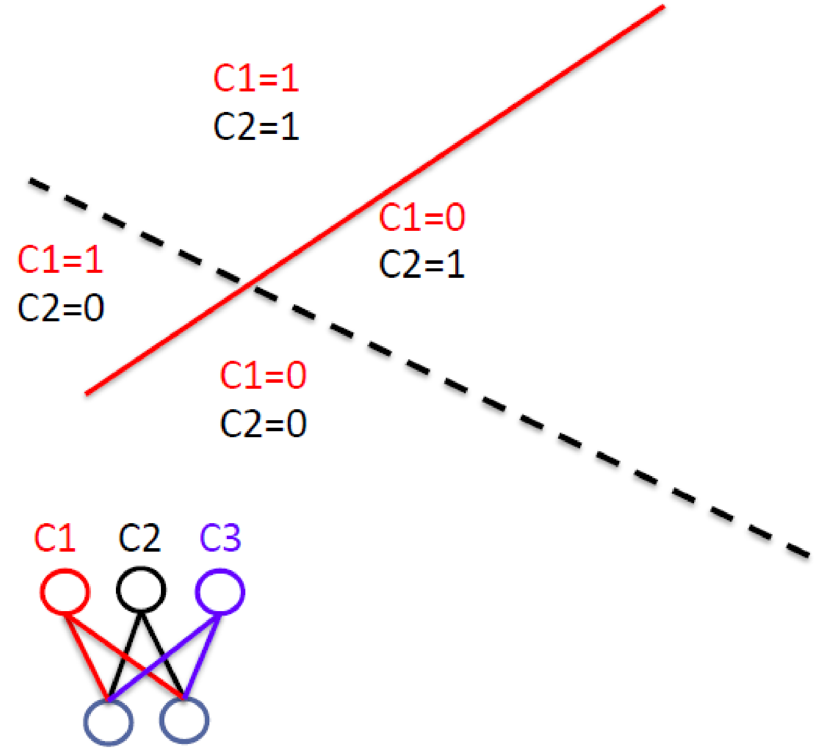
Two kinds of models

Nearest neighbors

- Look-up tables



Neural nets



Bengio, 2009, Foundations and Trends in Machine Learning

When to use machine learning

- Easy to get examples, hard to come up with an algorithm
- Don’t know how to program a solution (e.g. speech recognition, language processing, translation rely heavily on data)
- We need to automatically tune or adapt the solution to the user
- The solution changes over time
- Question: when shouldn’t we use learning?

Programming vs. learning

- Requires thorough problem understanding
- Formally define pre- and post-conditions
- Implement the solution
- Prove the correctness

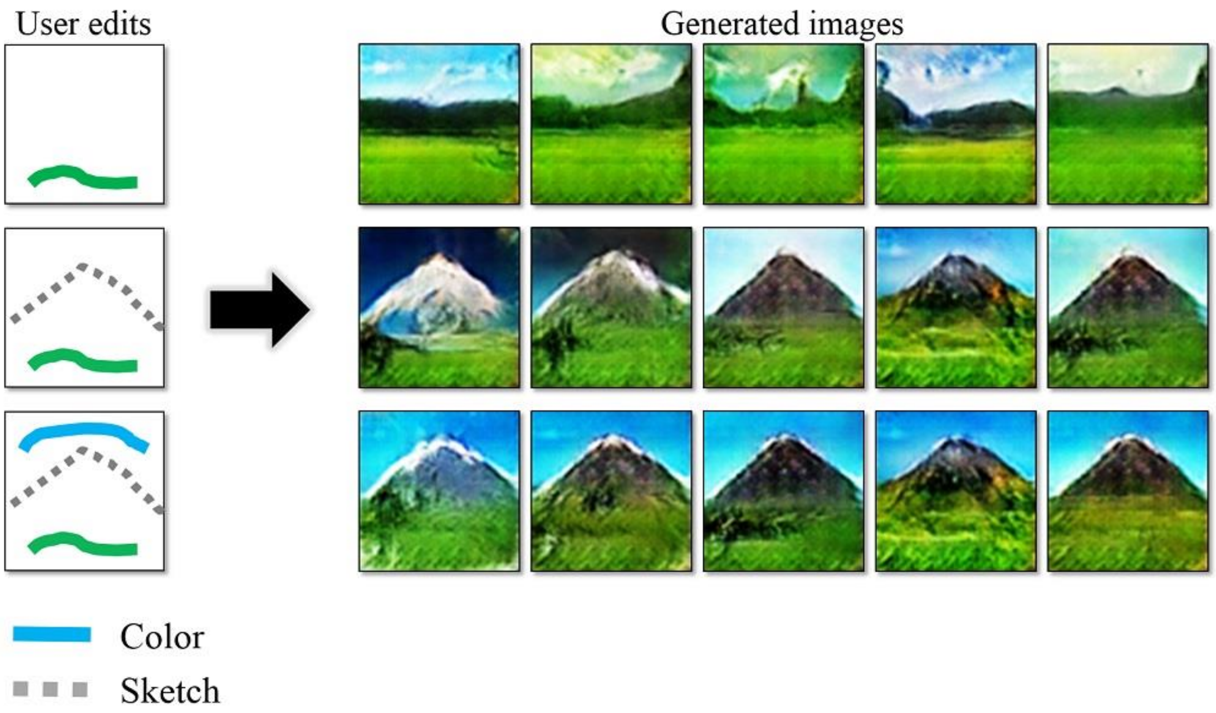
- Can have only a partial understanding of the problem – intuitions and a-priori assumptions
- Collect many examples of input-output
- Crucial aspect – **generalization** – will the learned solution work on new data?

Brainstorming ideas

What would you do with a good generative model of images?

Image manipulation

Transform sketches into images:



https://www.youtube.com/watch?v=FDELBFSeqQs

https://www.youtube.com/watch?v=9c4z6YsBGQ0

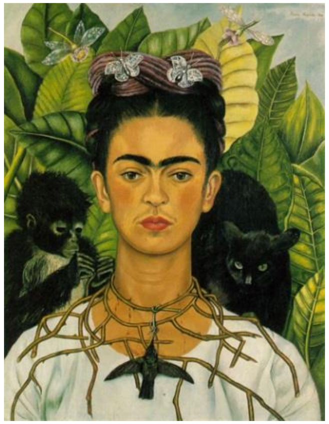
Image super-resolution

(e) Bicubic (f) SRCNN (g) A+ (h) RAISR

https://arxiv.org/pdf/1606.01299.pdf

Style transfer

+



=

Find image that takes content from image A and style from B

Gatys et al., „A Neural Algorithm of Artistic Style", 2015

Sample adaptation to videos:

https://www.youtube.com/watch?v=Khuj4ASldmU

Ruder et al., „Artistic Style Transfer For Videos"

Change the image to see many eyes/buildings in it.

Inceptionism: Going Deeper into Neural Networks
http://googleresearch.blogspot.com/2015/06/inceptionism-going-deeper-into-neural.html

Grocery Trip: https://www.youtube.com/watch?v=DgPaCWJL7XI

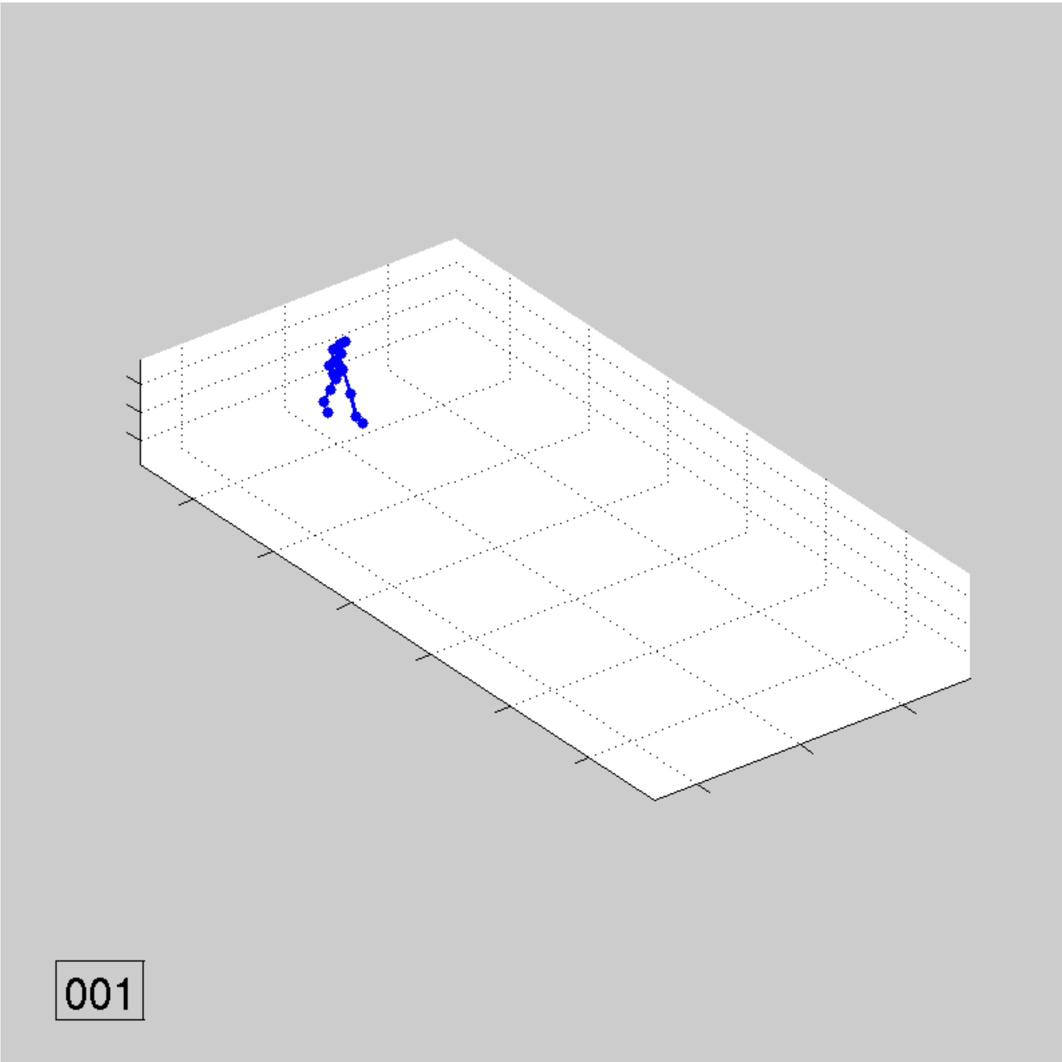
In between supervised and unsupervised learning

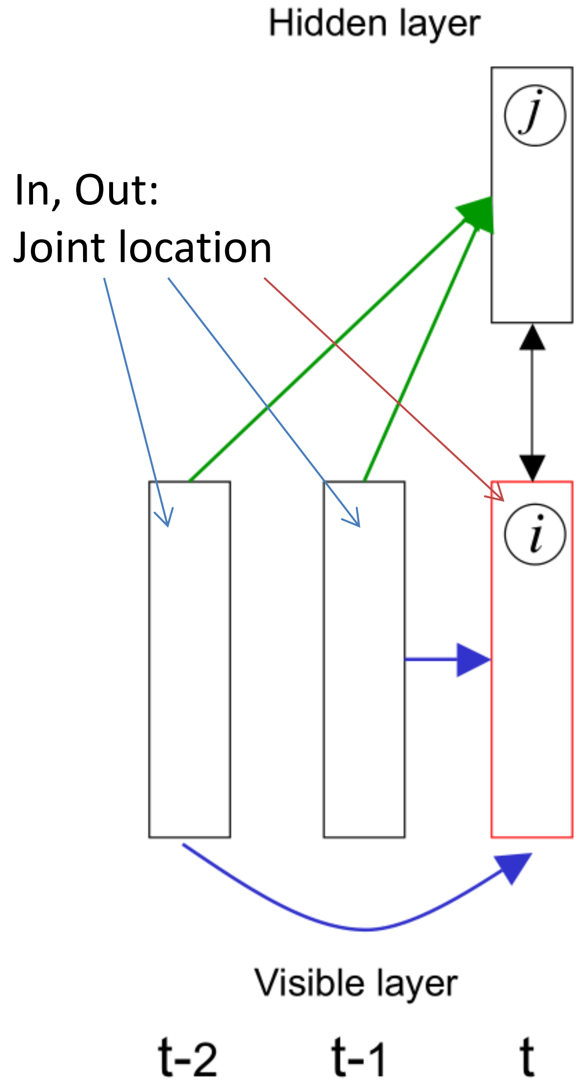
- Sequence generation
  - Inherent targets -> predict the next/neighboring elements
- Typical use: language modeling

  "Good everybody. Thank you very much. God bless the United States of America, and has already began with the world's gathering their health insurance. It's about hard-earned for our efforts that are not continued."

  (RNN generated Obama speech – by @samim)

Network to generate skeleton animations





Demo: http://www.uoguelph.ca/~gwtaylor/publications/nips2006mhmublv/

Reinforcement learning

- Learning to play games
  - Historically: backgammon
  - Now: Atari games https://www.youtube.com/watch?v=EfGD2qveGdQ
- Learning to perform movements (robotics) https://www.youtube.com/watch?v=EtMyH_--vnU&t=25m

Supervised and Reinforcement Mix

- AlphaGO:
  - 2 neural networks:
    - Position evaluator
    - Move proposer
  - First train to mimic human moves
  - Then do self-play and reinforcement learning