Machine learning for vision
Fall 2013
Roland Memisevic

Lecture 0, September 4, 2013

Objectives

- Learn about the recent advances in data driven vision.
- Learn how to apply some state-of-the-art learning and inference techniques in vision tasks.
- Learn about the basics and peculiarities of natural images statistics.
- (+ Get some ideas about visual information processing in biological systems.)

What is this course about

- Vision looks easy to humans.
- It is robust and flexible.
- It runs on fairly general-purpose hardware.

how many cars in the picture?
What is this course about

- Computer vision spent ≈ 50 years trying to mimic human vision.
- Huge inventory of tools: edge detectors, corner detectors, descriptors (e.g., SIFT), optic flow, hough transform, projective geometry...
- Unfortunately, it is difficult to make these work nicely together.

Roland Memisevic Machine learning for vision

A lower bound on the number of all images

- Assume your retina was only 16 × 16 pixels large and you could see only black and white.
- There are still $2^{16 \times 16} = 2^{256}$ possible images.
- So there are more tiny binary images than there are atoms in the universe.
- And even more large color images.

Roland Memisevic Machine learning for vision

An upper bound on the number of images you will see in your life

- Assume you see 100 images per second, 3600 seconds per hour, 24 hours per day.
- This is < 10 mio images per day, or 3.65 billion images per year.
- So you will see < 300 billion images in your life and you had seen < 10 billion images when you turned 3.
- This is a tiny number compared to the number of possible images.
- Yet, at that age you were a champion at recognizing and reasoning about unfamiliar objects.
Natural images are not random

- As compared to the number of possible images, there is a diminishingly small number of natural images!

Roland Memisevic Machine learning for vision

Random images

Natural images (berkeley database)

Natural images (grayscale)
The distribution over natural images has low entropy.

If images are “random”, you will need 256 bits on average to transmit each.

If your images are structured, you will need much fewer bits.

For example, what if the images contain two square blocks of random size at random locations?

(Hyvarinen et al, 2009)
You can transmit the upper-left corner and the bottom-right corner each with 8 bits, making it $2 \times 16 = 32$ bits for both squares.

(It could be more efficient than that.)

Caveat: Neural codes, ironically, are very high-dimensional. It is the entropy of each individual code element that is small. This leads to sparse representations.

Another way to state that the information content is small is to say that there are dependencies among the pixels.

A common way to reduce the dependencies is Independent Components Analysis (ICA)
View from neuroscience

- Attneave 1954, Barlow 1961

What do visual neurons like to see?

- Hubel and Wiesel, 1959

A very simple neuron abstraction

\[ y = w^T x \]

Unsupervised Learning

\[ y = W^T x \]
Learning criteria

▶ maximize independence (ica)
▶ minimize entropy (information theory)
▶ maximize sparseness (sparse coding)
▶ maximize probability of the data (eg. boltzmann machines, mixture models)
▶ learn to reconstruct from bottleneck (contractive autoencoder)
▶ supervised learning (eg. learn to classify objects)

Learned receptive fields

CIFAR challenge

(Krizhevsky, et al. 2009)

eg. Coates et al., 2011
ImageNet challenge

▶ Convolutional nets: (Hubel and Wiesel, 1959); Fukushima, 1980; LeCun et al., 1990; Riesenhuber and Poggio 1990
▶ Vision solved?
Vision is more than object recognition

how many cars in the picture?

Roland Memisevic Machine learning for vision

There are things images can’t teach you

Roland Memisevic Machine learning for vision
There are things images can’t teach you

(Buelthoff and Buelthoff, 2003)

how many chairs in the picture?
A slightly more general “neuron”

- Processing videos works better with a different kind of neuron abstraction (Hinton, 1981), (Adelson and Bergen, 1985)
- Similarly for 3D vision, geometry, invariant recognition, tracking, etc.

Activity recognition example

- Convolutional GBM (Taylor et al., 2010)
- hierarchical ISA (Le, et al., 2011)

Watching videos improves recognition


Contour coding

- (Hyvarinen et al., 2002)
Tracking with mocap

(Taylor, et al.; 2010)

Major conferences and journals

- **NIPS**: Neural Information Processing Systems
- **CVPR**: International Conference on Computer Vision and Pattern Recognition
- **ICCV**: International Conference on Computer Vision
- **ICML**: International Conference on Machine Learning
- **ECCV**: European Conference on Computer Vision
- **PAMI**: IEEE Transactions on Pattern Analysis and Machine Intelligence
- **Neural Computation**
- **JMLR**: Journal of Machine Learning Research

Course outline

1. Fourier representations and Gabor features
2. Basic image statistics, aspects of biological vision
3. Feature learning
4. Energy models, motion, invariance
5. Advanced topics and applications
Advanced topics

- Saliency and attention.
- Fixations, part-based models.
- Tracking.
- Scene understanding.
- Activity understanding.
- Attributes.
- Detachable objects.
- Multi-modality.

Unofficial textbook


Learning approach

- Readings will be posted and should be read before each class.
- Lectures will explain and motivate the concepts with real world examples.
- Student presentations of recent papers to discuss recent/novel/speculative/applied ideas.
- Several hands-on assignments to get an idea for how the methods work on actual data.
- Final projects are research based. Eg., evaluation/comparison of an approach from a recent paper, prototype/discussion of a new idea or variation of an existing one.

Marking scheme

- readings (10 %)
- participation in class (20 %)
- assignments (30 %)
- term project (40 %)
Relation to other courses and areas

- **Image Processing, Computer Vision**: Focus on *data and learning* (and *bio-inspired* as a consequence).
- **Neuroscience**: The brain (and neuroscience) is utterly complex and detailed. We will abstract away a *lot* of these details.
- **Machine Learning**: Images have *strong structure*. Black-box classifiers (like SVM) and fully Bayesian / variational methods not always the best choice.

Tuesday

- Review of linear algebra, stats, optimization, complex arithmetic, ...