

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 this course is about



how many cars in the picture?

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

What this course is about



how many cars in the picture?

Roland Memisevic

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

What this course is about



how many cars in the picture?

- Huge progress in recent years, based on a single simple idea:
- Images are not random.
 → Treat vision as a statistical inference task.

▲■▶ ▲≣▶ ▲≣▶ = ● のQC

◆□▶ ◆□▶ ◆ □▶ ◆ □▶ ● □ ● ● ●

Roland Memisevic Machine learning for vision

A lower bound on the number of all images 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.</p>
- This number is tiny compared to the number of possible images.
- Yet, at that age you were a champion at recognizing and reasoning about unfamiliar objects.
- The number of *labeled* images is of course much smaller yet.



 Assume your retina was only 16 × 16 pixels large and you could see only black and white.

Machine learning for vision

There are still 2^{16*16} = 2²⁵⁶ possible images.

(=115792089237316195423570985008687907853269984665640564039 457584007913129639936)

- So there are more tiny binary images than there are atoms in the universe.
- And even more large color images.

・ロ・・一部・・ヨ・・ヨ・ シュウ





Roland Memisevic Machine learning for vision

View from information theory



- You can transmit the upper-left corner and the bottom-right corner each with 8 bits (4 for the vertical, 4 for the horizontal direction), making it 2 × 16 = 32 bits for both squares.
- (It could be more efficient than that.)

Machine learning for vision

◆□ → ◆□ → ◆ 三 → ◆ 三 → ○ へ ⊙

View from information theory



 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.

▲■▶ ▲≣▶ ▲≣▶ = ● のQC

View from statistics

Roland Memisevic



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



Roland Memisevic



 A common way to reduce the dependencies is *Independent Components Analysis* (ICA)

◆□ ▶ ◆□ ▶ ◆三 ▶ ◆□ ▶ ◆□ ● ● ●

Machine learning for visior

▲□▶▲圖▶▲≣▶▲≣▶ = ● のQ@



Learning criteria

- maximize independence (ica)
- minimize entropy (information theory)
- maximize sparseness (sparse coding)

Roland Memisevic

- maximize probability of the data (eg. boltzmann machines, mixture models)
- learn to reconstruct from bottleneck (autoencoders, kmeans)

Machine learning for vision

◆□▶ ◆□▶ ◆三▶ ◆三▶ ◆□▶

supervised learning (eg. learn to classify objects)

Learned receptive fields





learned receptive fields

real receptive fields

Roland Memisevic Machine learning for vision

ImageNet challenge

	SuperVision	test-preds-141-146.2009-131- 137-145-146.2011-145f.	0.15315	Using extra training data from ImageNet Fall 2011 release	
	SuperVision	test-preds-131-137-145-135- 145f.txt	0.16422	Using only supplied training data.	
	ISI	pred_FVs_wLACs_weighted.txt	0.26172	Weighted sum of scores from each classifier with SIFT+FV, LBP+FV, GIST+FV, and CSIFT+FV, respectively.	
	151	pred_FVs_weighted.txt	0.26602	Weighted sum of scores from classifiers using each FV.	
	ISI	pred_FVs_summed.txt	0.26646	Naive sum of scores from classifiers using each FV.	
	ISI	pred_FVs_wLACs_summed.txt	0.26952	Naive sum of scores from each classifier with SIFT+FV, LBP+FV, GIST+FV, and CSIFT+FV, respectively.	
				Mixed selection from High-Level SVM scores	
Krizhe	vsky, et	al. 2012			

ImageNet challenge





<section-header><section-header><section-header><section-header><section-header><section-header><section-header><section-header><section-header><image><image><image>

There are things images can't teach you

Roland Memisevic Machine learning for vision

There are things (still) images can't teach you

how many chairs in the picture?

Machine learning for vision

(Buelthoff and Buelthoff, 2003)

・ロト・西ト・ヨト・ヨー もくの

∃ <\0</p>

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

Tracking with mocap

Machine learning for vision

(Taylor, et al.; 2010)

▲□▶▲□▶▲≧▶▲≧▶ ≧ めぬの

Course outline

1. Fourier representations and Gabor features

Roland Memisevic

- 2. Basic image statistics, aspects of biological vision
- 3. Feature learning
- 4. Energy models, motion, invariance
- 5. Advanced topics, applications, recent work

Machine learning for vision

◆□ > ◆□ > ◆三 > ◆三 > ● ● ● ●

