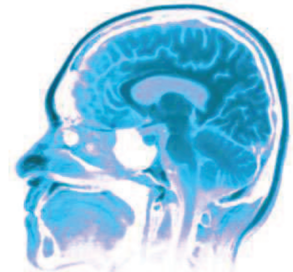
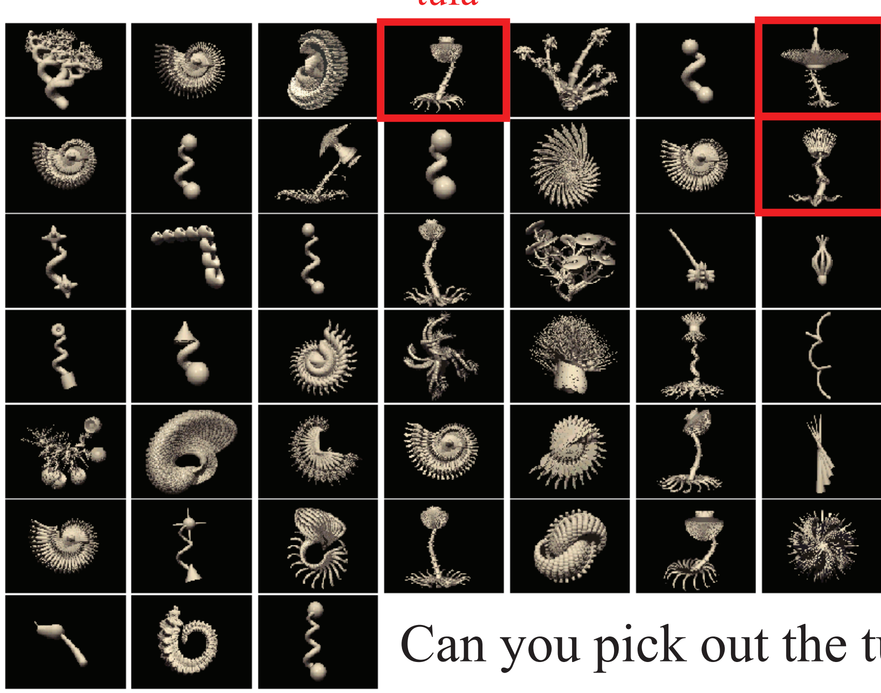
CPSC540



Introduction to Machine Learning
And Neural Computation

Nando de Freitas
September, 2011
University of British Columbia

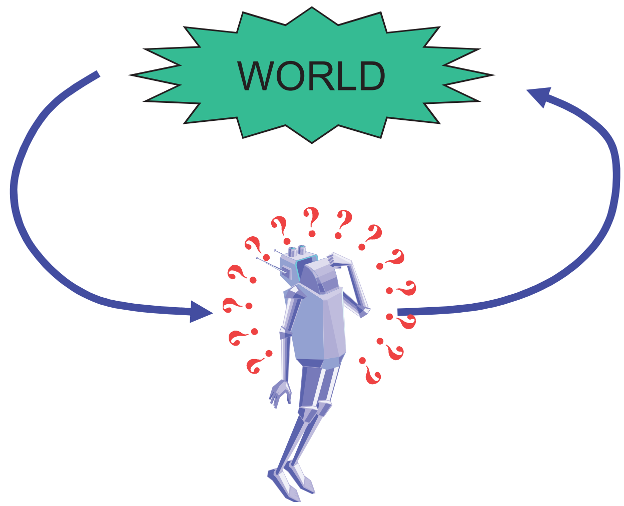


Can you pick out the tufas?

Josh Tenenbaum

Learning

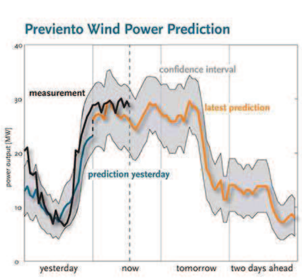
``Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the task or tasks drawn from the same population more efficiently and more effectively the next time."
-- Herbert Simon

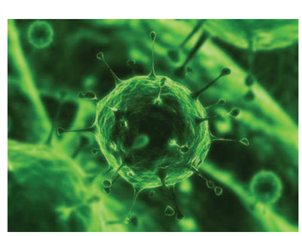


Machine Learning

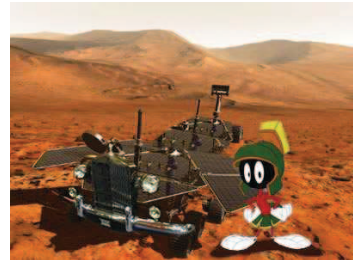
Machine learning deals with the problem of extracting features from data so as to solve predictive tasks:

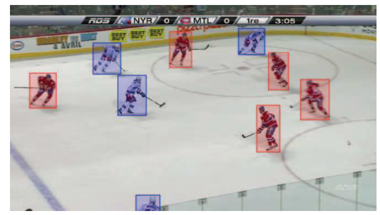
- ❑Forecasting (*e.g. Energy prediction*)
- ❑Imputing missing data (*e.g. Netflix recommendations*)
- ❑Detecting anomalies (*e.g. Intruder detection, disease control*)
- ❑Classifying (*e.g. Credit risk assessment, diagnosis*)
- ❑Ranking (*e.g. Google search*)
- ❑Summarizing (*e.g. News aggregators*)
- ❑Decision making (*e.g. AI agents*) …



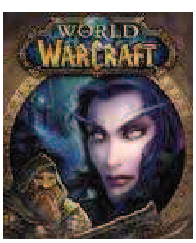


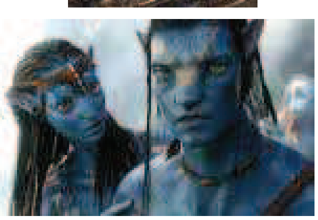
When to apply machine learning





- Human expertise is absent (*e.g.* *Navigating on Mars*)
- Humans are unable to explain their expertise (*e.g.* *Speech recognition, vision, language*)
- Solution changes with time (*e.g.* *Tracking, temperature control*)
- Solution needs to be adapted to particular cases (*e.g.* *Biometrics*)
- The problem size is to vast for our limited reasoning capabilities (*e.g.* *Calculating webpage ranks and matching ads to facebook pages*)





- **Library of Congress** text database of ~20 TB
- **AT&T** 323 TB, 1.9 trillion phone call records.
- **World of Warcraft** utilizes 1.3 PB of storage to maintain its game.
- **Avatar** movie reported to have taken over 1 PB of local storage at *Weta Digital* for the rendering of the 3D CGI effects.
- **Google** processes ~24 PB of data per day.
- **YouTube:** 24 hours of video uploaded every minute. More video is uploaded in 60 days than all 3 major US networks created in 60 years. According to *cisco*, internet video will generate over 18 EB of traffic per month in 2013.

Machine learning in NLP

"Large" text dataset:

- **1,000,000 words in 1967**
- **1,000,000,000,000 words in 2006**

Success stories:

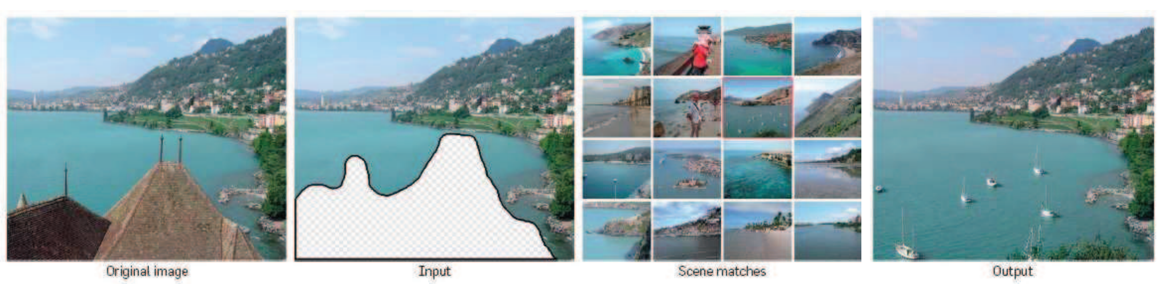
- **Speech recognition**
- **Machine translation**

What is the common thing that makes both of these work well?

- **Lots of labeled data**
- **Memorization is a good policy**

[Halevy, Norvig & Pereira, 2009]

Scene completion: more data is better



Given an input image with a missing region,
Efros uses matching scenes from a large
collection of photographs to complete the image

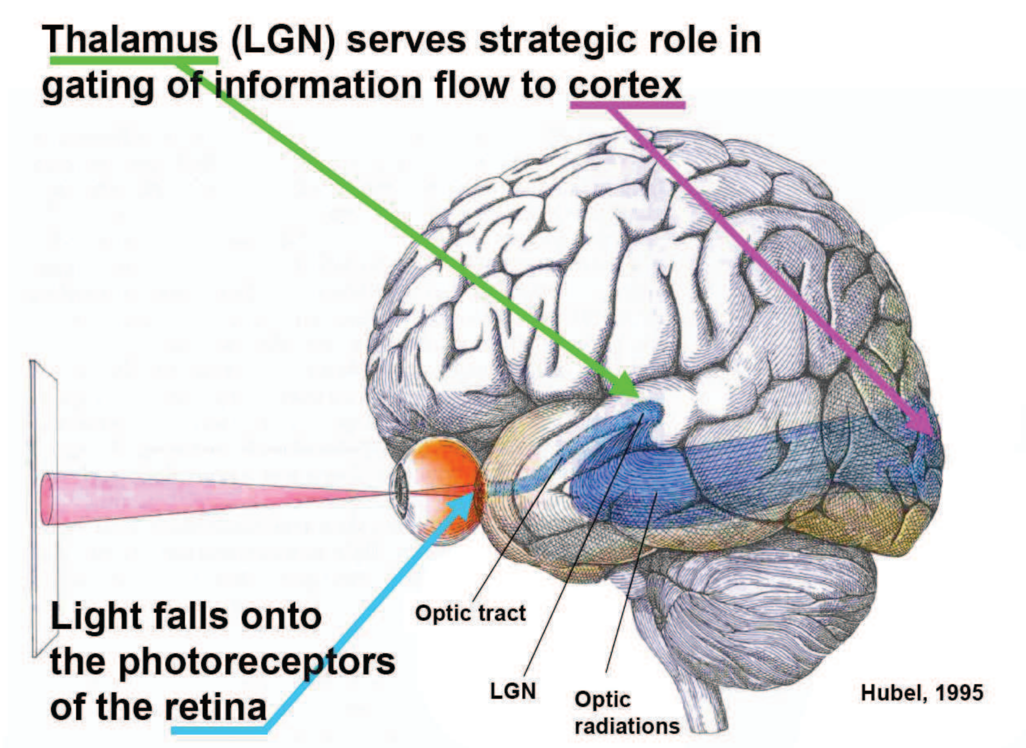
[Efros, 2008]

The semantic challenge

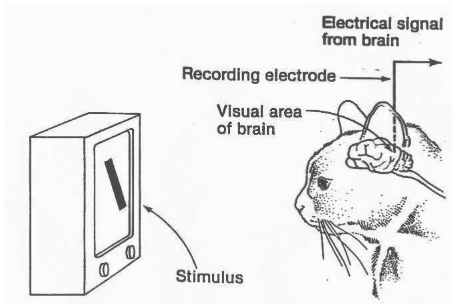
- “We’ve already solved the sociological problem of building a network infrastructure that has encouraged hundreds of millions of authors to share a trillion pages of content.
- We’ve solved the technological problem of aggregating and indexing all this content.
- But we’re left with a scientific problem of interpreting the content”
- **It’s not only about how big your data is. It is about understanding it and using this understanding to derive reasonable inferences. Think of citation matching.**

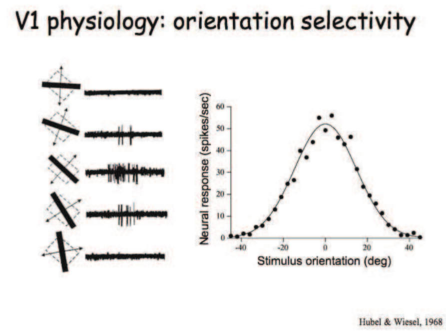
[Halevy, Norvig & Pereira, 2009]

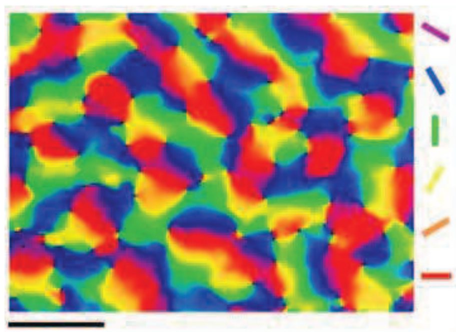
A source of inspiration

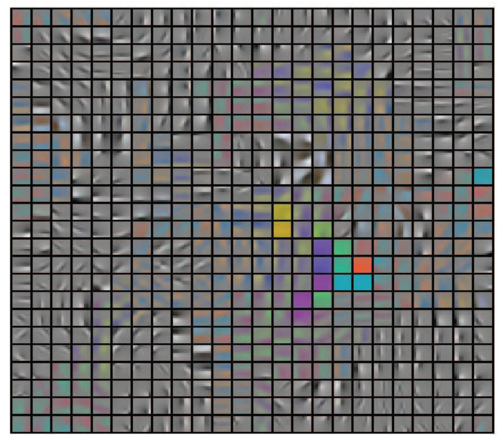


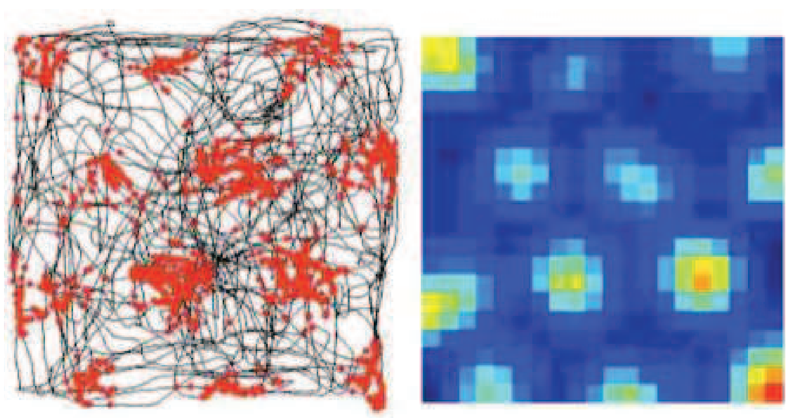
Selectivity and Topographic maps in V1







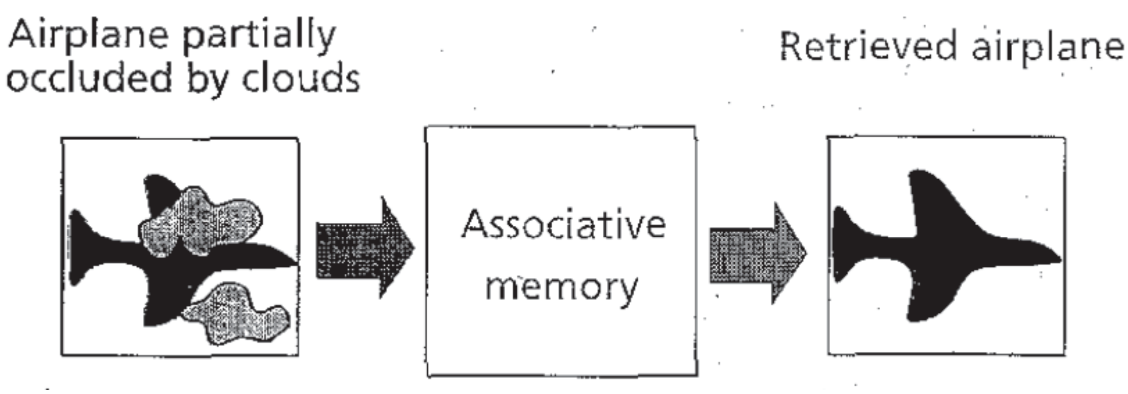




"the x and y coordinates correspond to the spatial location of a rat, which is running around freely inside a large box. The black lines in the left figure shows how this particular rat explored the box in a fairly haphazard manner. However, an electrode inserted in the rat's subcortex picks up a signal that is anything but chaotic: the responses of said neuron are given as red dots in the left figure, while the right figure gives the firing rate distribution (ranging from blue for silent and red for the peak rate of responding). Although the rat is running about randomly, this neuron is responding in a grid, seemingly coming on an off in response to the animal's spatial location."

[Hafting et al 2005]

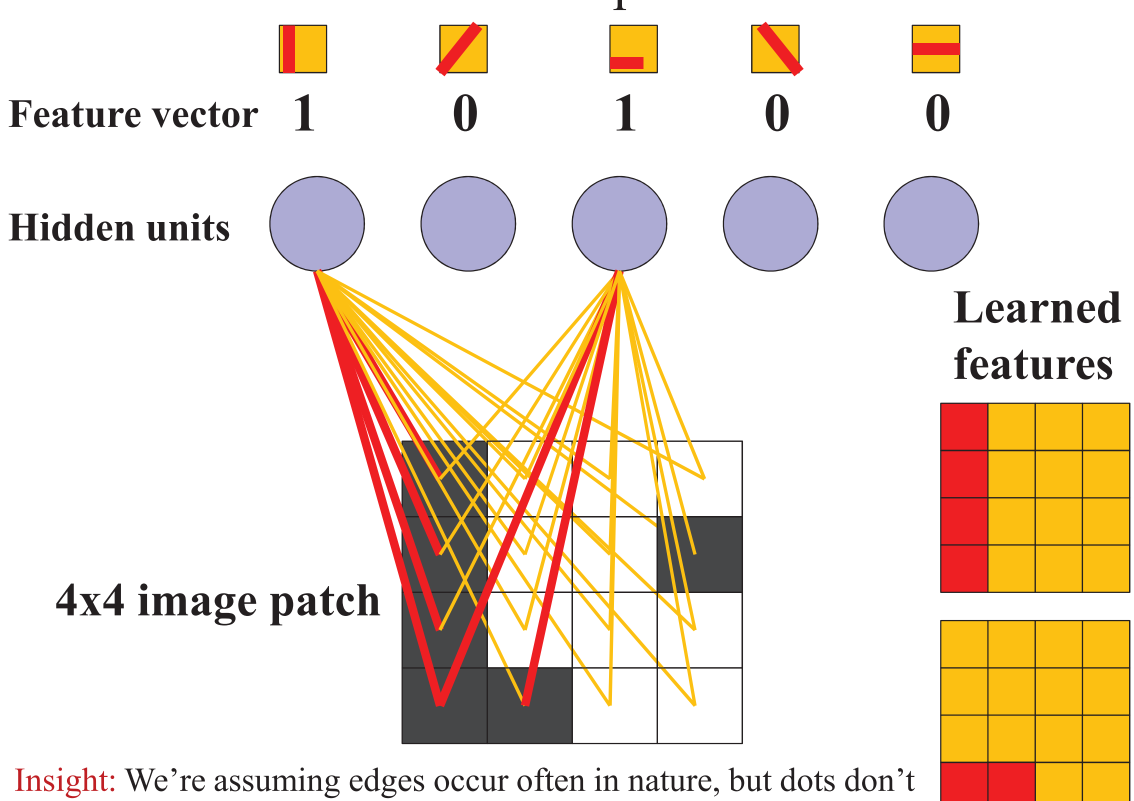
Associative memory



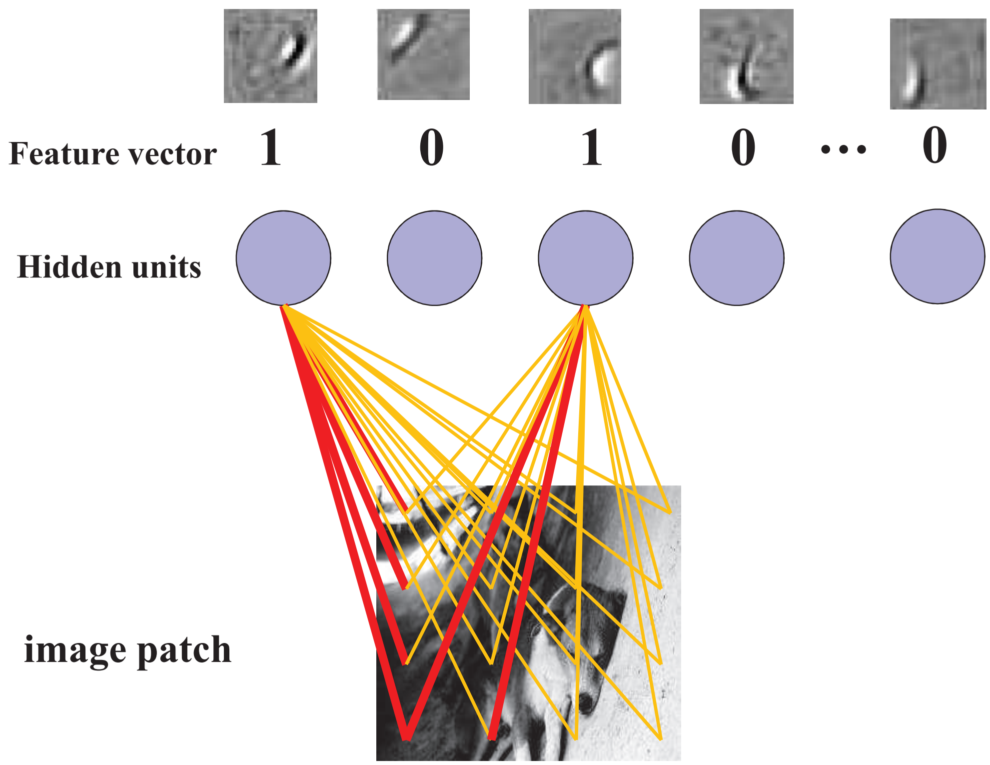
Example 2: Say the alphabet, …. backward

[Jain, Mao & Mohiuddin, 1996]

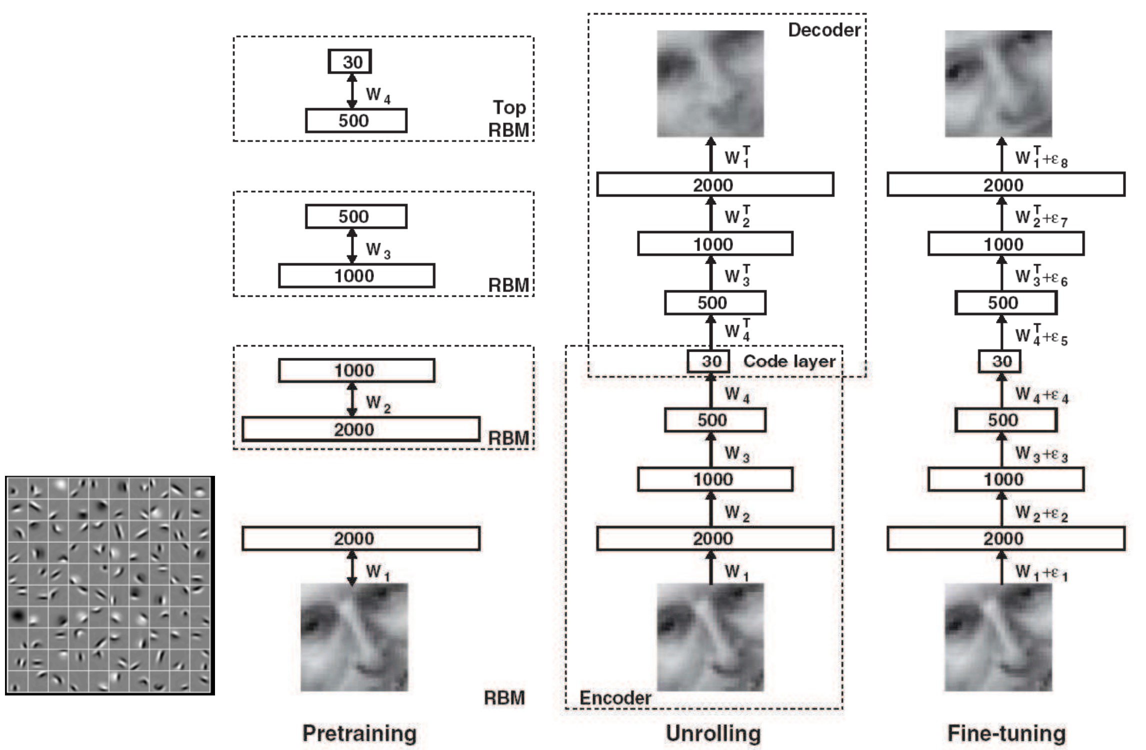
Distributed representation



Insight: We're assuming edges occur often in nature, but dots don't
We learn the regular structures in the world



Deep learning (Hinton and collaborators)

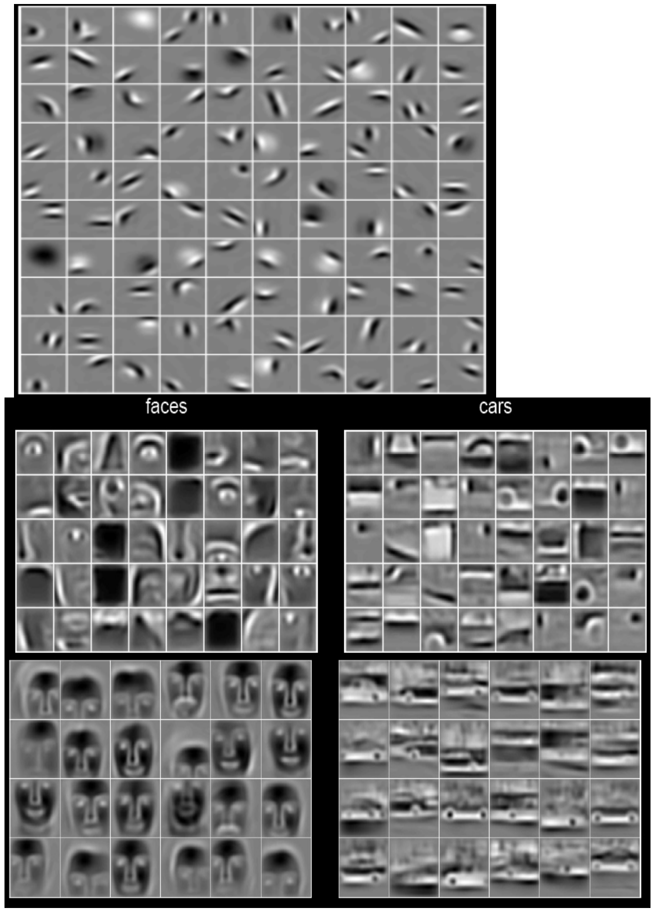


[Russ Salakhutdinov and Geoff Hinton]

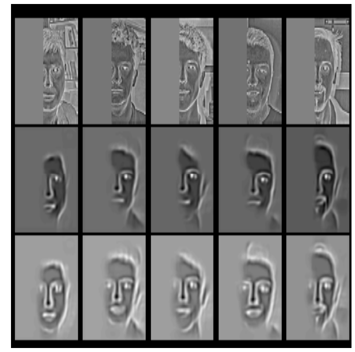
Layer 1

Layer 2

Layer 3



Completing scenes



[Honglak Lee et al 2009]

Hierarchical spatio-temporal feature learning

Observed gaze sequence

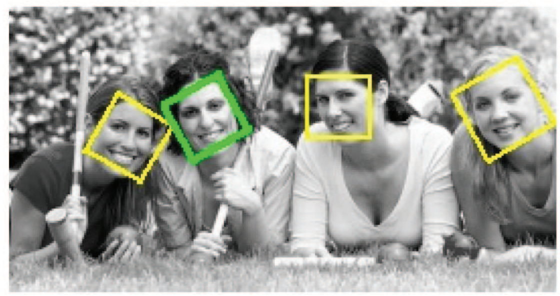
Model predictions

[Bo Chen et al 2010]

Face detection

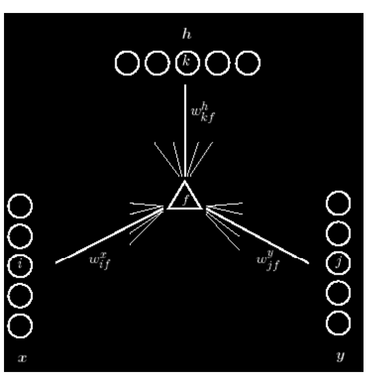
We can use the feature activations to predict whether an image patch has a face.

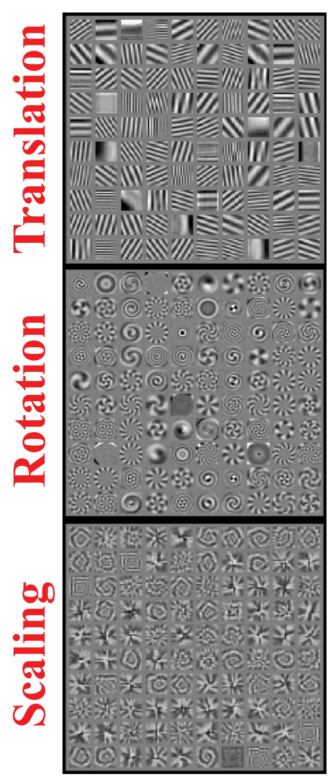
- Divide the image into many small overlapping patches at different locations, scales and orientations, and classify each such patch based on whether it contains face-like texture or not. This is called a **sliding window detector**.



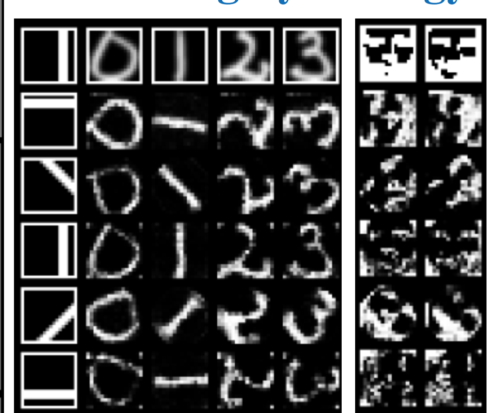
[Kevin Murphy 2010]

Learning image transformations and analogy



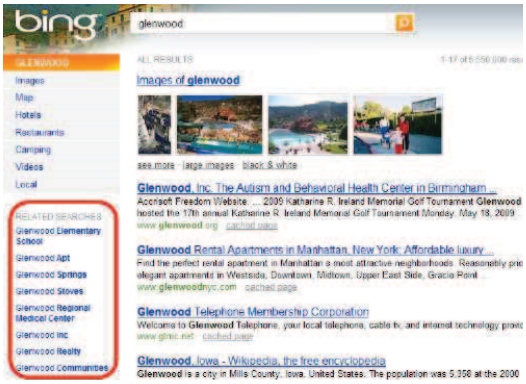


Learning by analogy



[Memisevic and Hinton 2009]

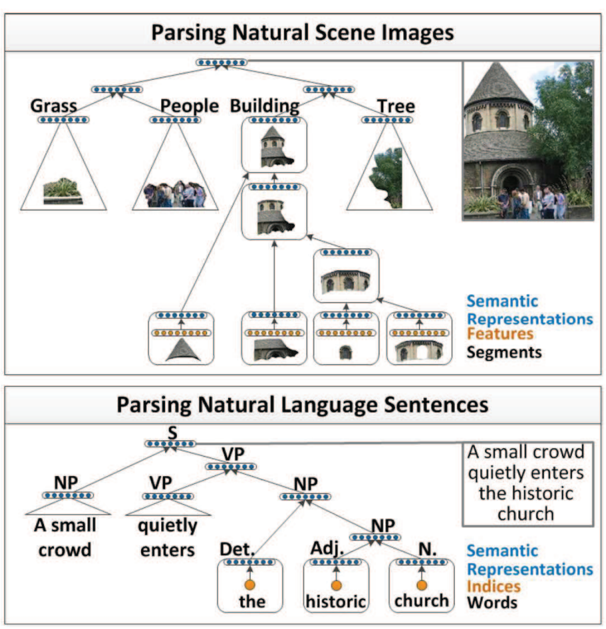
Relational learning

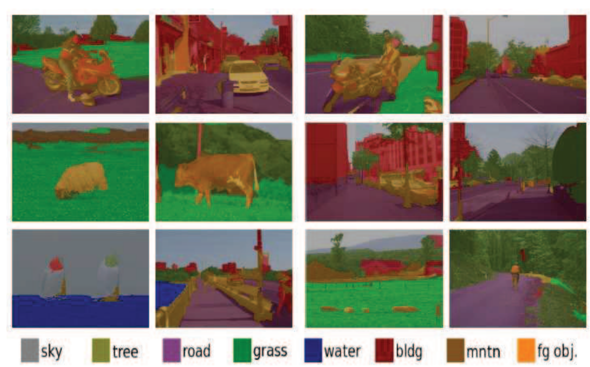


e^l	people				
r	build	destroy	won	suffer	control
e^r	*livelihoods*	*icons*	*emmy*	*sores*	*rocket*
	homes	*virtue*	*award*	*agitation*	*stores*
	altars	*donkeys*	*everything*	*treatise*	*emotions*
	houses	*cowboy*	*standings*	*eczema*	*spending*
	ramps	*chimpanzees*	*pounds*	*copd*	*fertility*

[Yoshua Bengio, Jason Weston et al 2011]

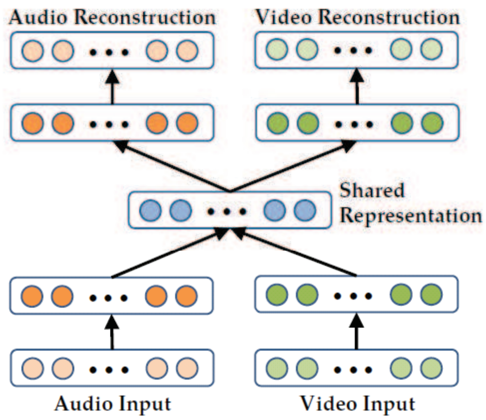
Learning to parse scenes and language





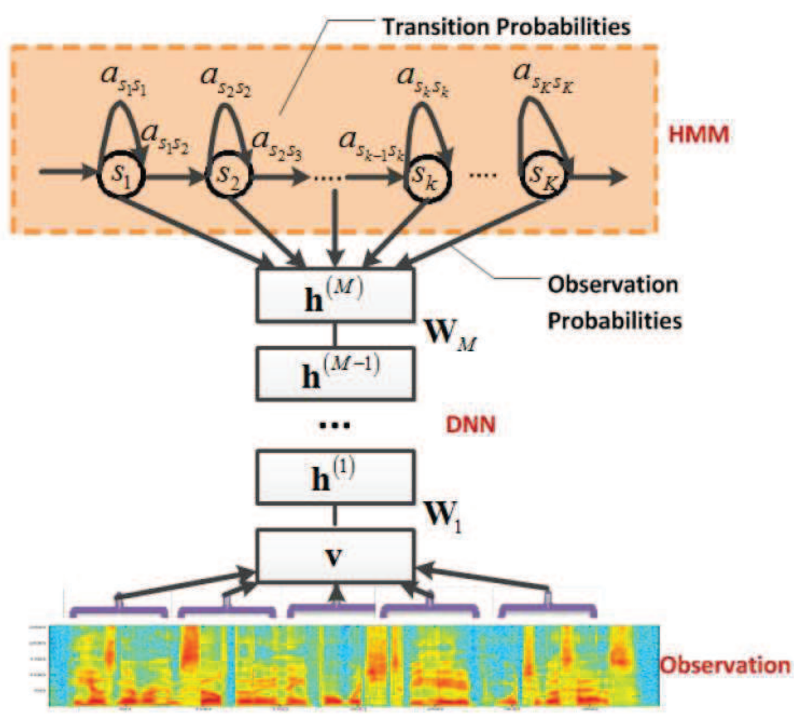
[Richard Socher et al 2011]

Multimodal autoencoders



[Jiquan Ngiam et al 2011]

Learning in speech recognition



[George Dahl et al 2011]