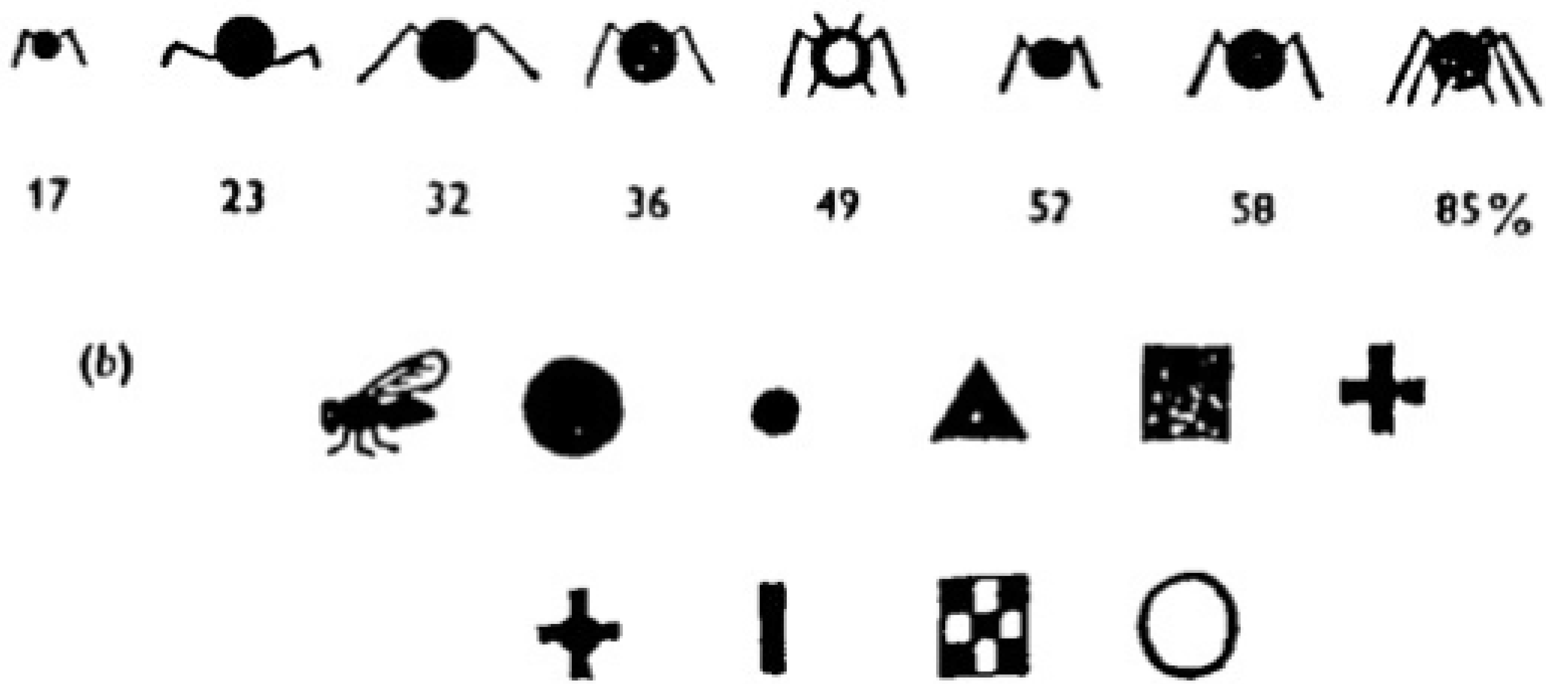
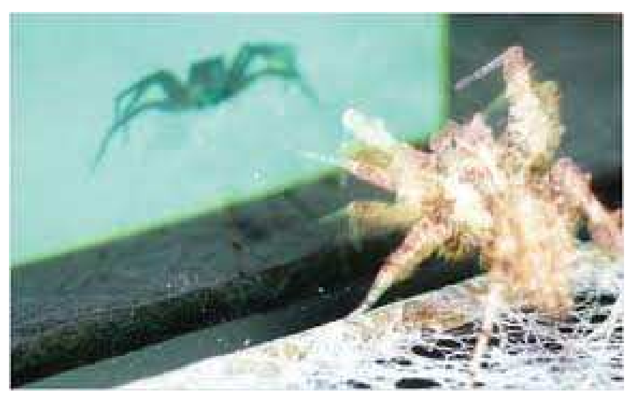
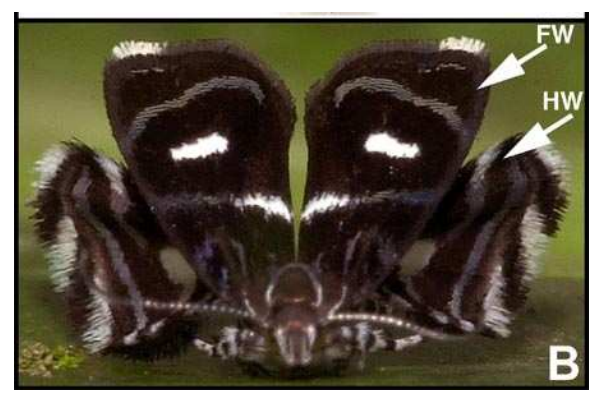
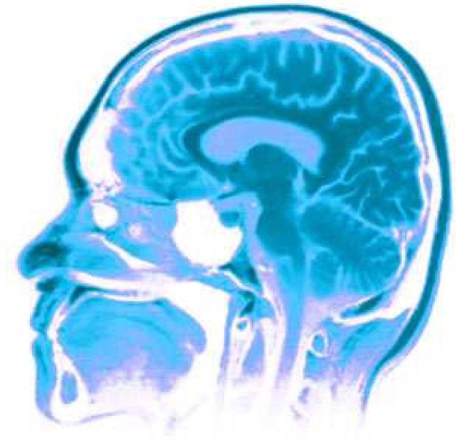


Text-fig. 12. Stimuli found by Drees to evoke courtship (*a*) and prey capture (*b*) in male jumping spiders (*Epiblemum scenicum*). The numbers beneath each figure in (*a*) are the percentage of trials on which courtship was evoked. After Drees (1952).





CPSC340



Machine Learning & Data Mining

Nando de Freitas

September, 2012

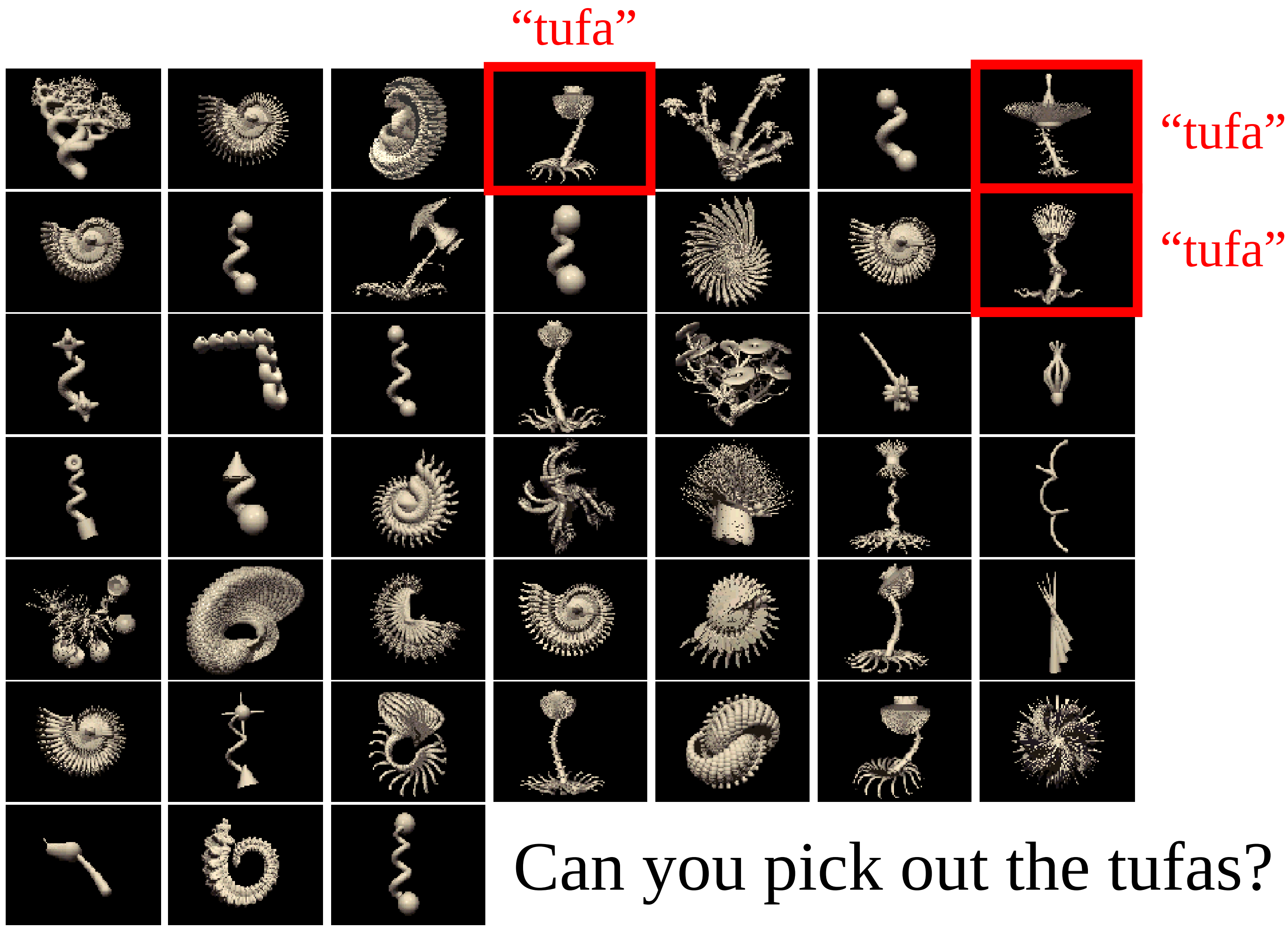
University of British Columbia

Outline of the lecture

This lecture provides an introduction to the course. It covers the following four areas:

1. **Definitions** of machine learning and data mining
2. The **big data** phenomenon
3. Drawing inspiration from **neural** systems
4. Machine learning **applications** and impact

The intent of the lecture is not to explain details of building ML systems, or to tell you what to study for the exam. Rather it is an overview of what can be accomplished with ML. If it inspires you, then you'll have to take the course and learn a lot of cool math in the process!



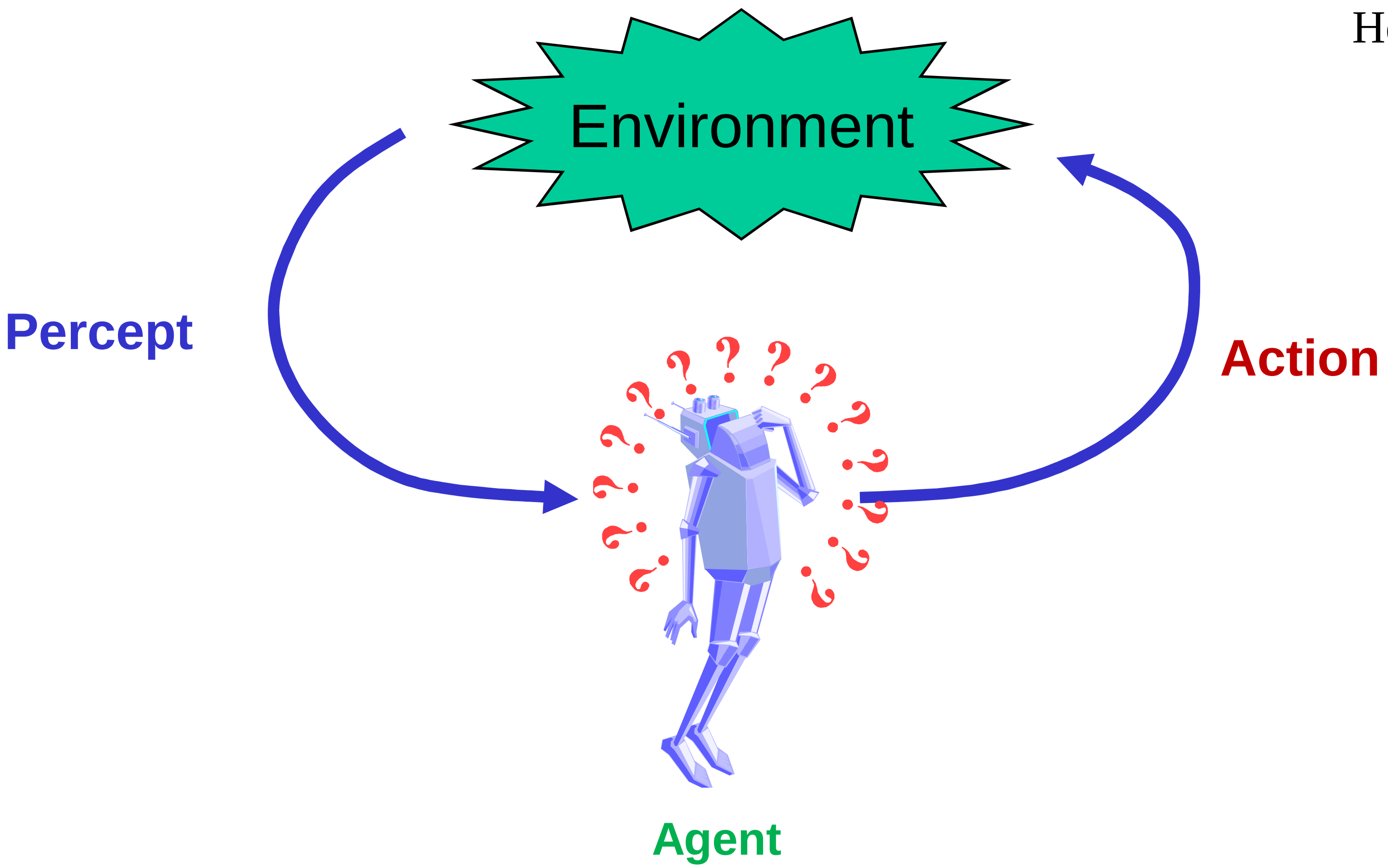
Can you pick out the tufas?

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.''*

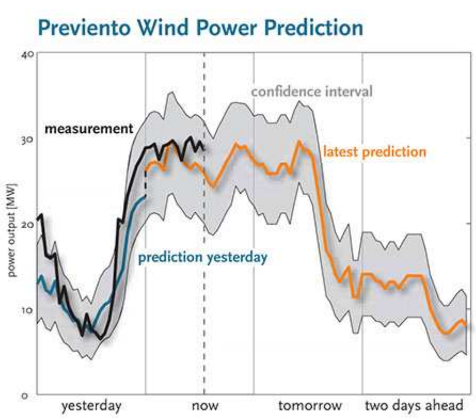
Herbert Simon

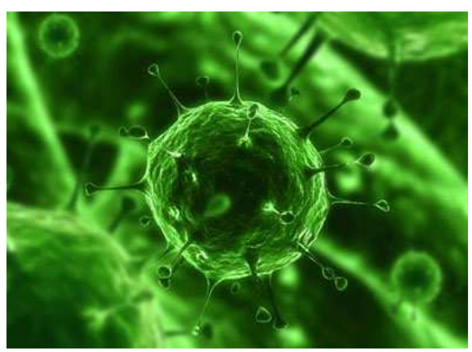


Machine learning

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

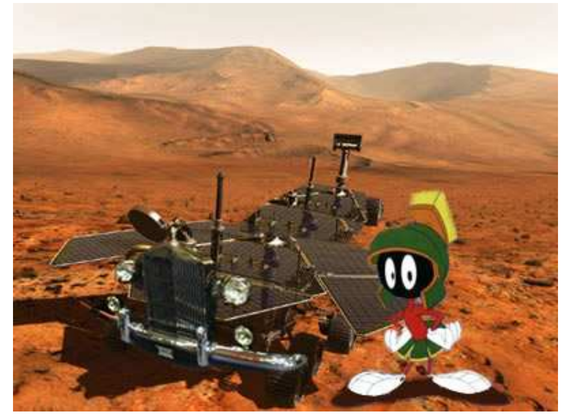
- Forecasting (*e.g. Energy demand prediction, sales*)
- Imputing missing data (*e.g. Netflix recommendations*)
- Detecting anomalies (*e.g. Intruders, virus mutations*)
- Classifying (*e.g. Credit risk assessment, cancer diagnosis*)
- Ranking (*e.g. Google search, personalization*)
- Summarizing (*e.g. News zeitgeist, social media sentiment*)
- Decision making (*e.g. AI, robotics, compiler tuning, trading*)

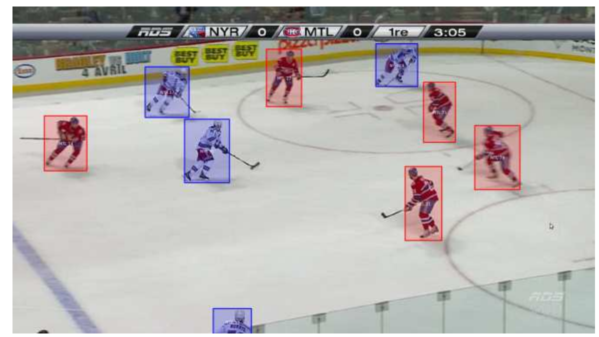




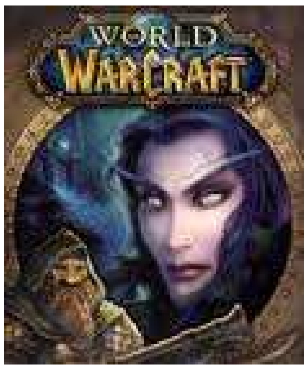

When to apply machine learning

- 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, preferences*)
- Solution needs to be adapted to particular cases (*e.g. Biometrics, personalization*)
- The problem size is to vast for our limited reasoning capabilities (*e.g. Calculating webpage ranks, matching ads to facebook pages*)





Big Data!



- **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 language

"Large" text dataset:

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

Success stories:

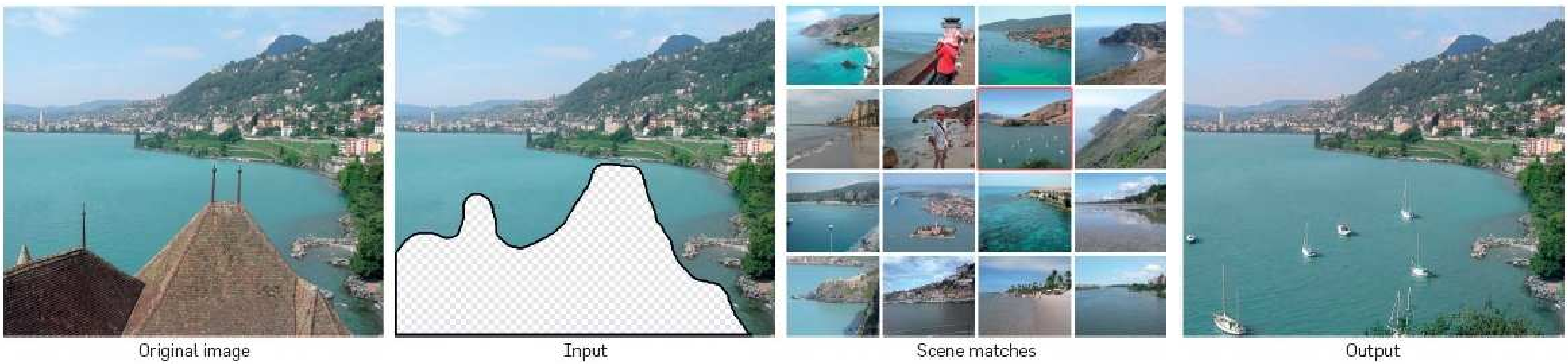
- **Speech recognition**
- **Machine translation**

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.**

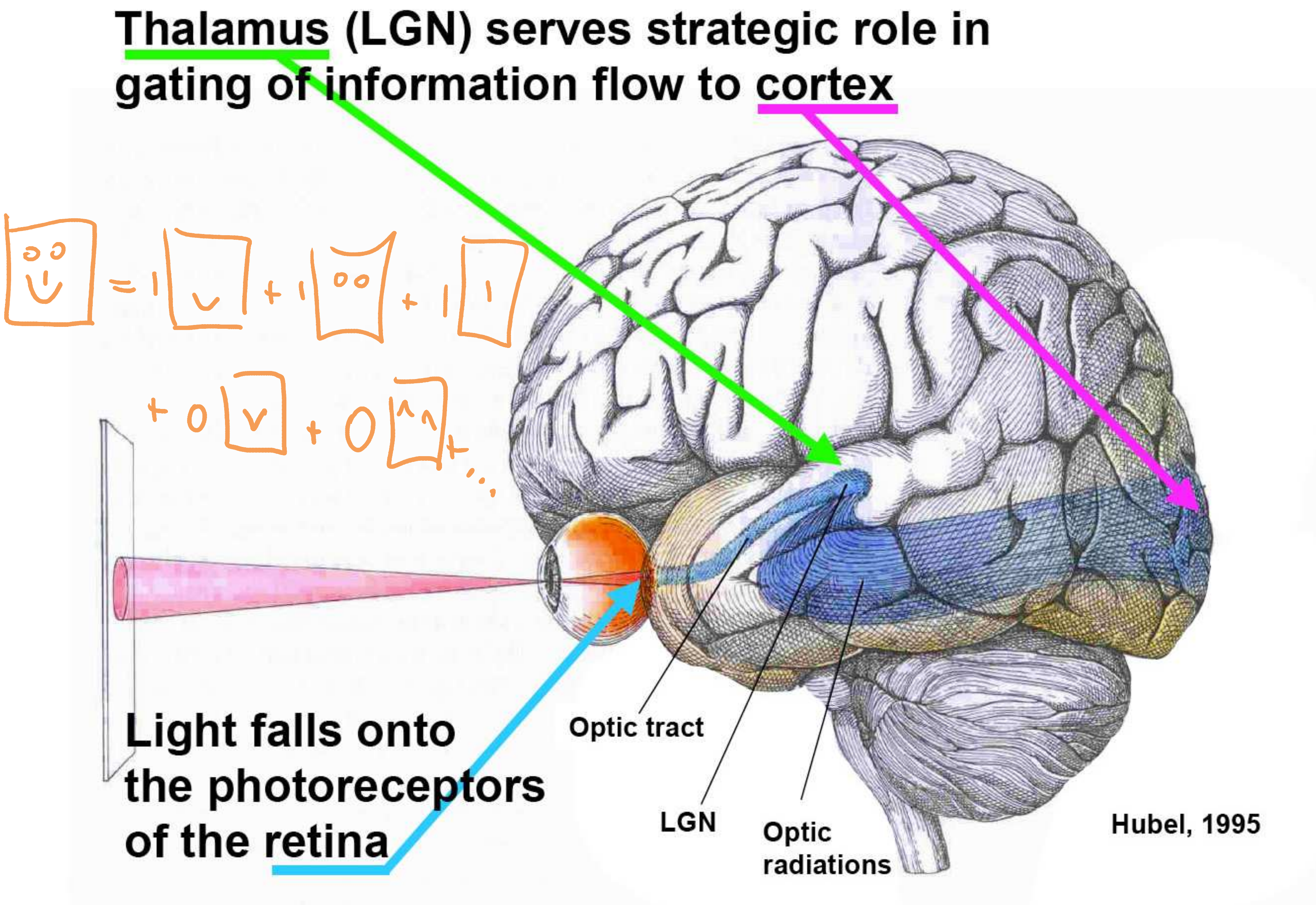
- **We’ve solved the technological problem of aggregating and indexing all this content.**

- **But we’re left with a scientific problem of interpreting the content”**

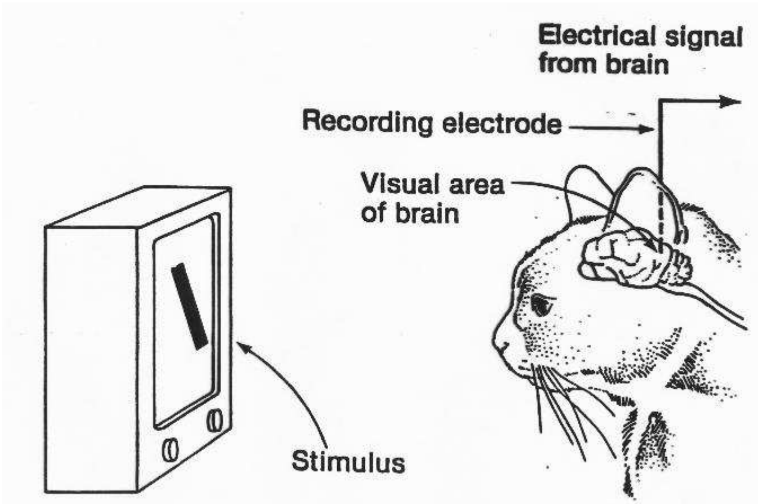
- **It’s not only about how big your data is. It is about understanding it and using this understanding to derive reasonable inferences. Think of citation matching.**

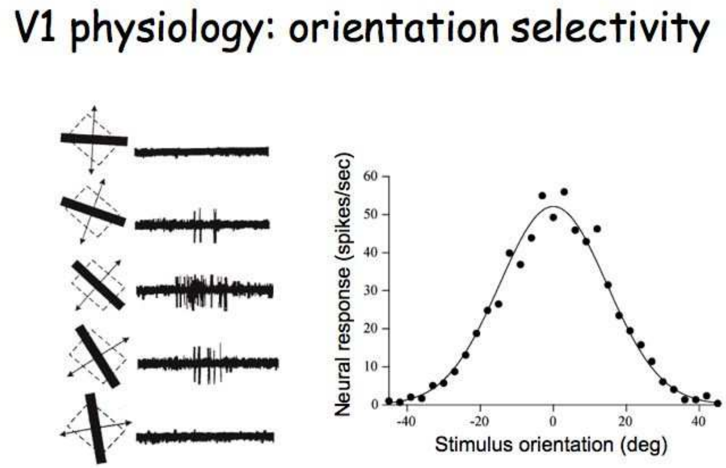
[Halevy, Norvig & Pereira, 2009]

A source of inspiration

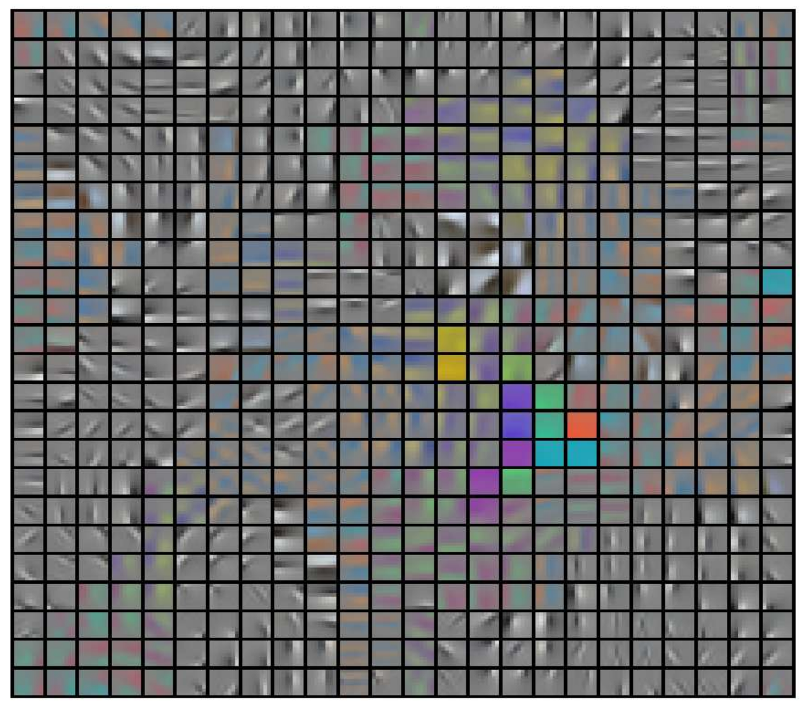


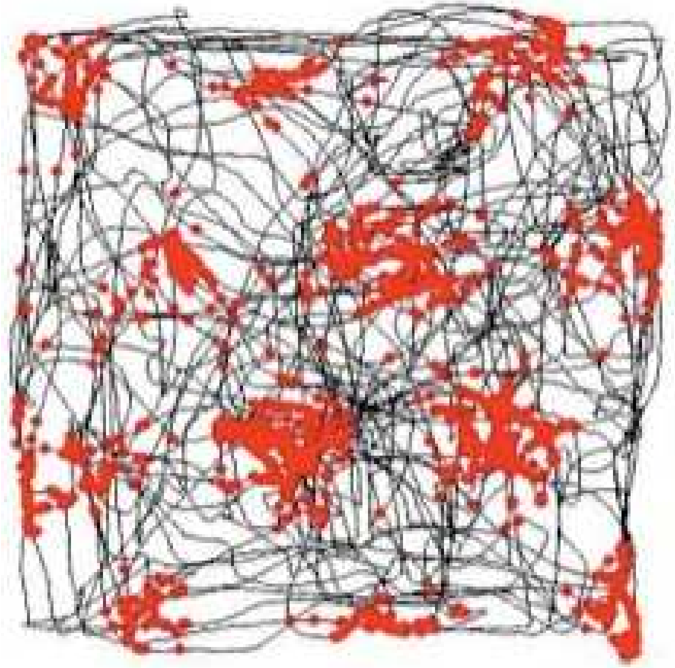
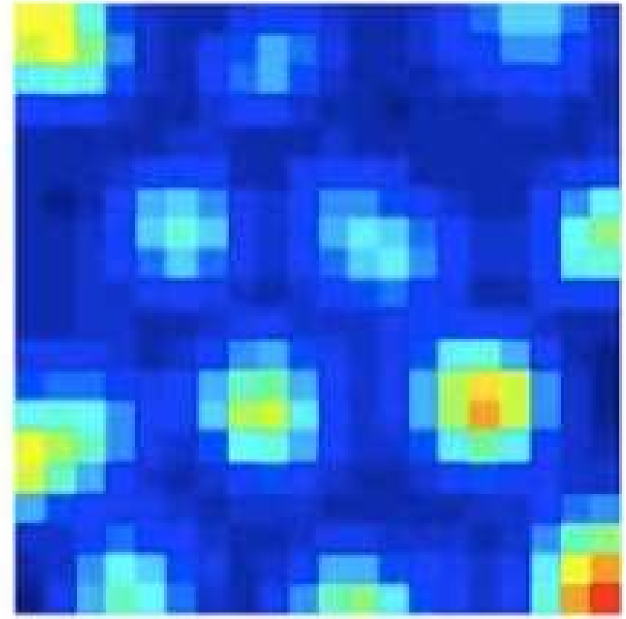
Selectivity and Topographic maps in V1





Hubel & Wiesel, 1968



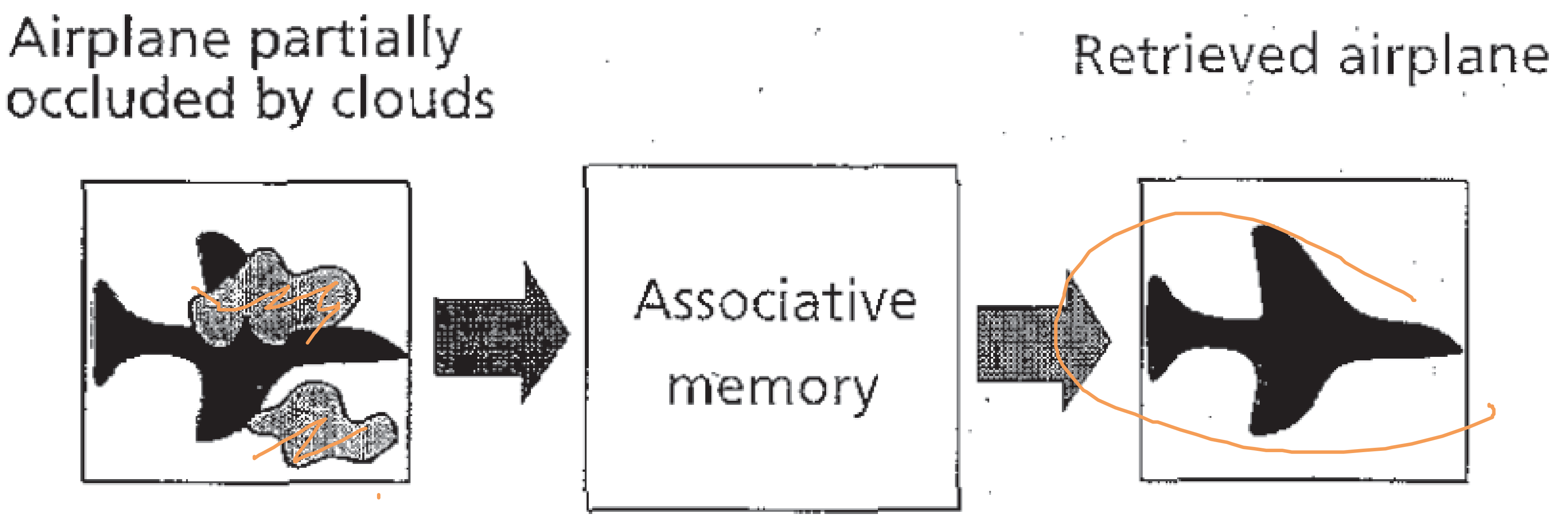




The x and y coordinates correspond to the spatial location of a rat.

The red dots indicate the place where a particular neuron fires.

[Hafting et al 2005]

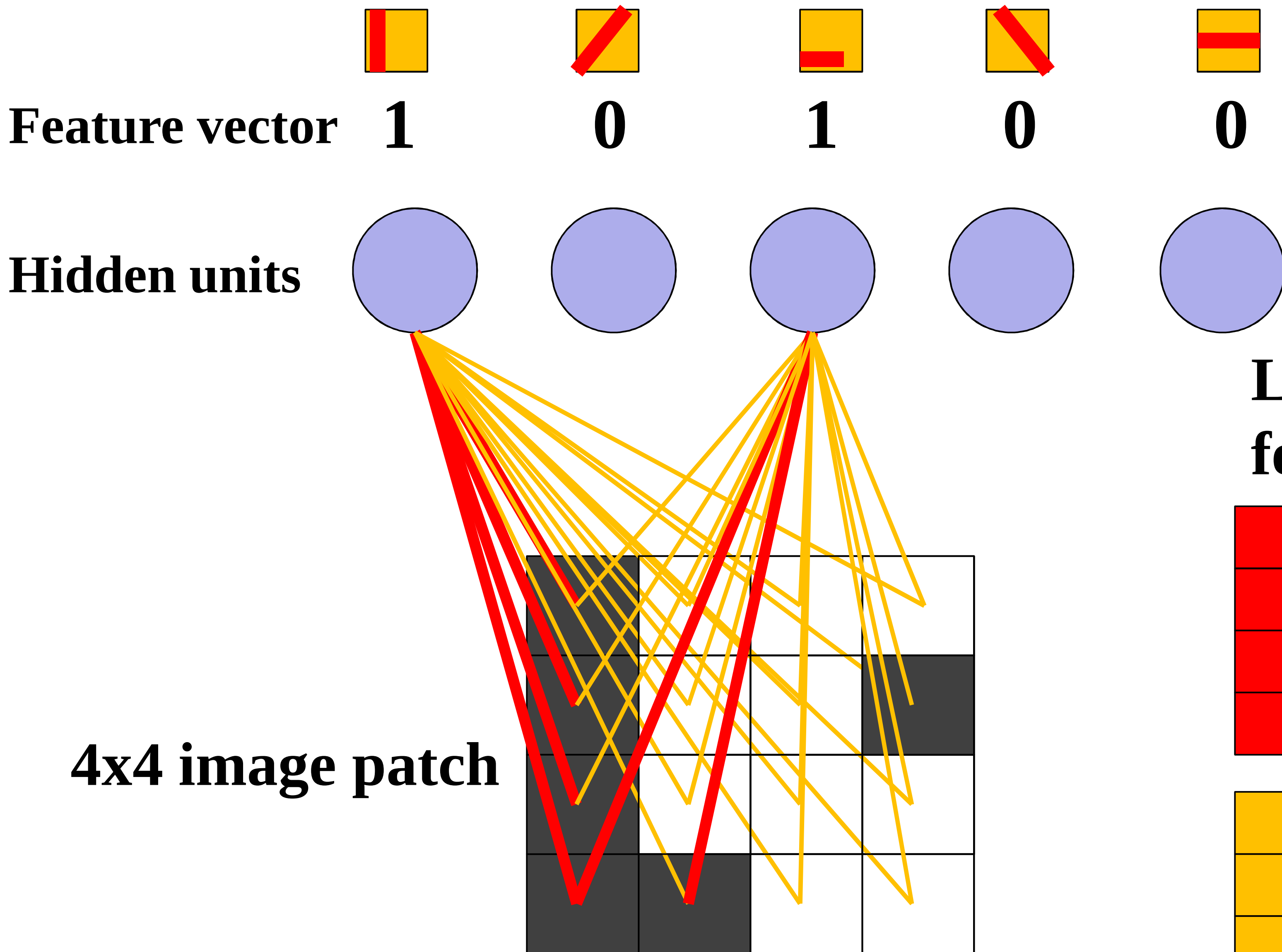
Associative memory

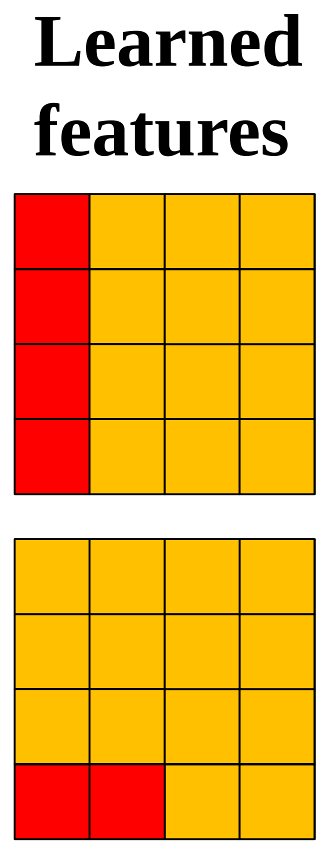


Example 2: Say the alphabet, …. backward

[Jain, Mao & Mohiuddin, 1996]

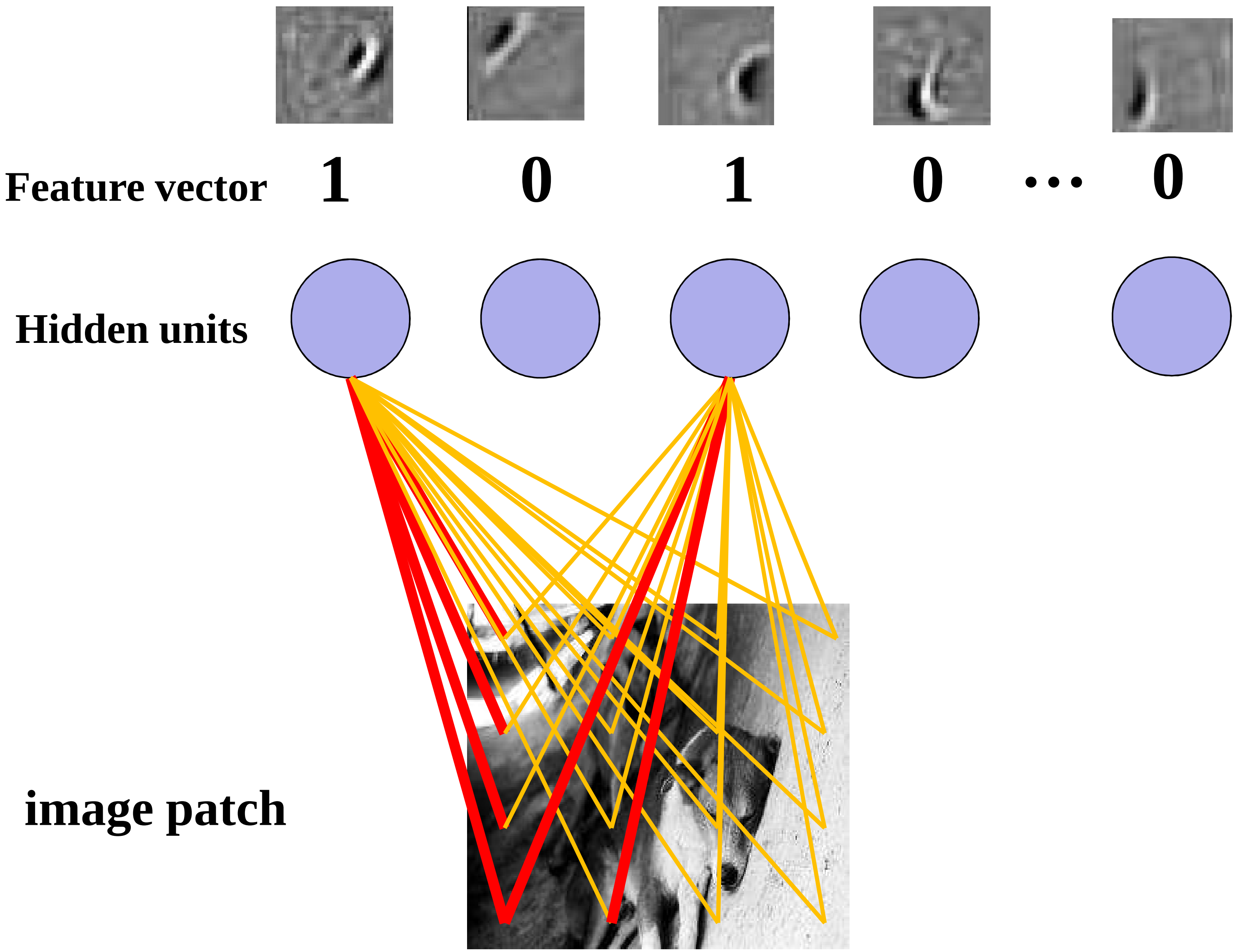
Neural network: A distributed representation



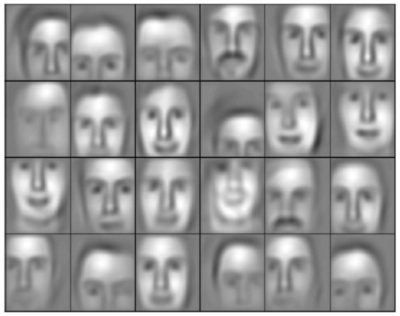


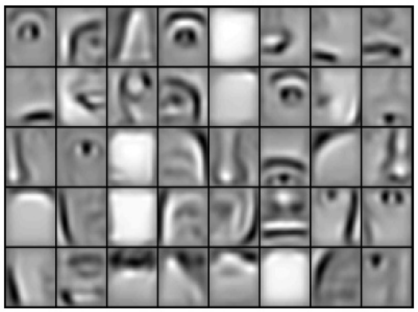
Insight: We’re assuming edges occur often in nature, but dots don’t

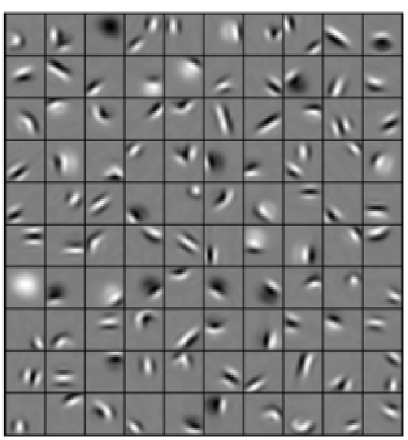
We learn the regular structures in the world

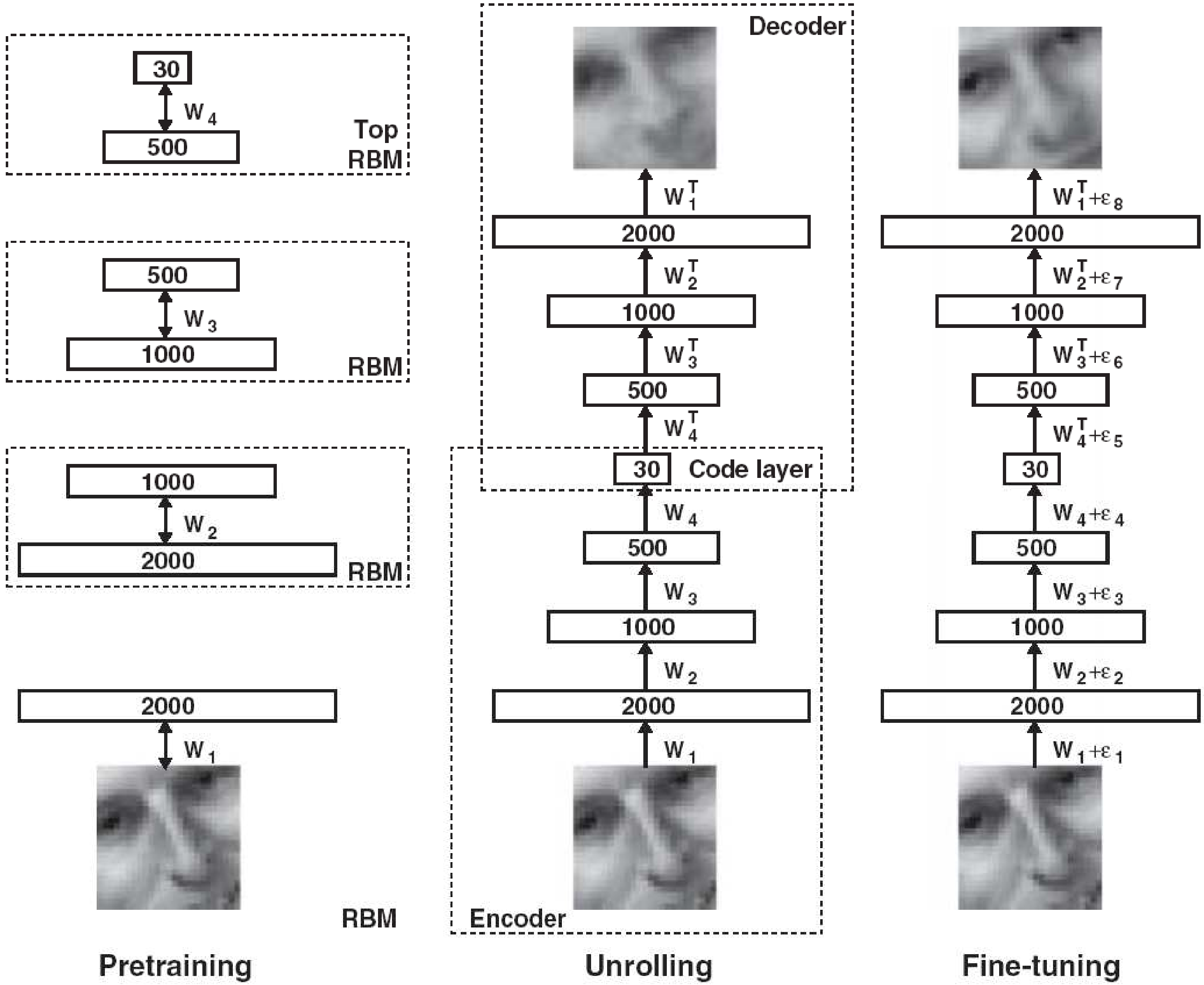

Feature vector
1
0
1
0
…
0
Hidden units
image patch

Deep learning with autoencoders



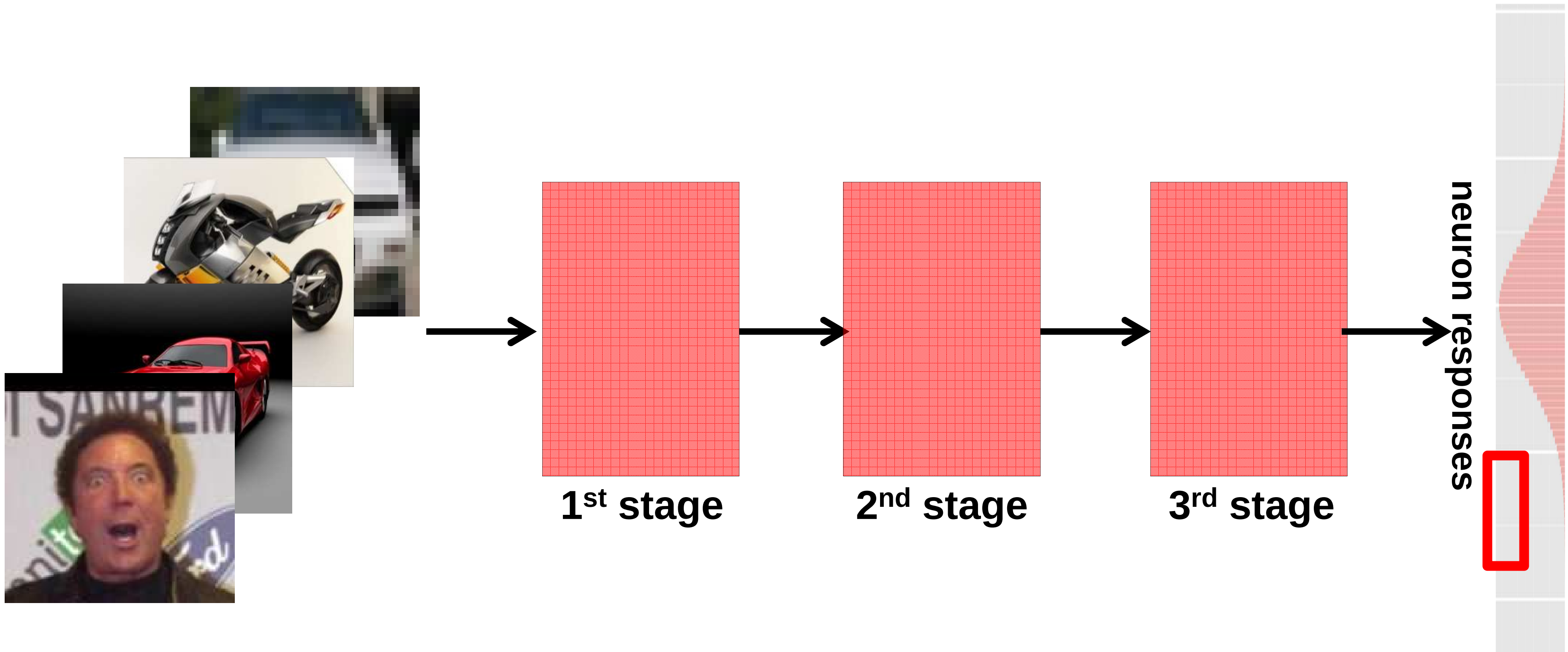






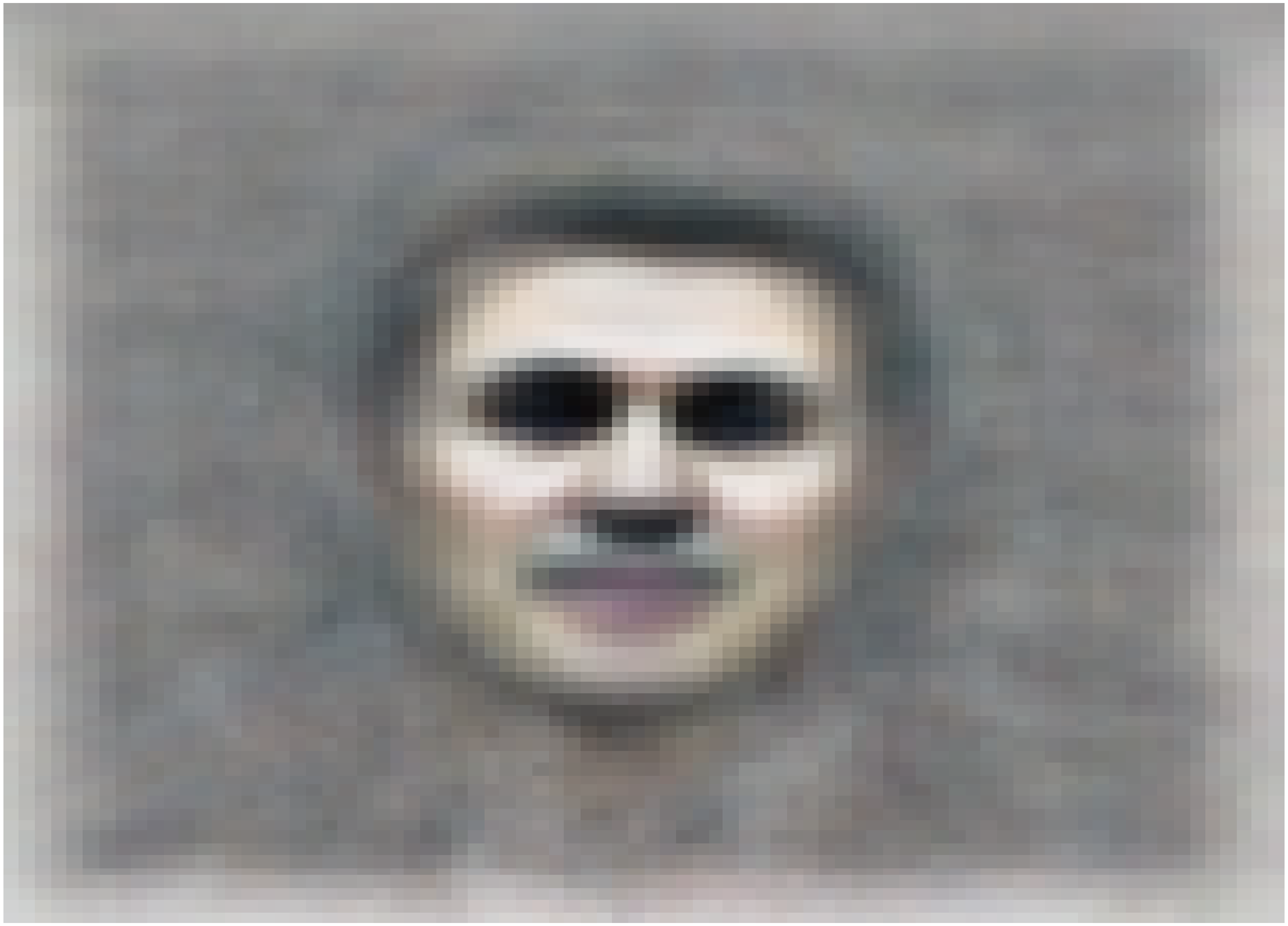
[Russ Salakhutdinov, Geoff Hinton, Yann Lecun, Yoshua Bengio, Andrew Ng, …]

Validating Unsupervised Learning



Top Images For Best Face Neuron

Best Input For Face Neuron



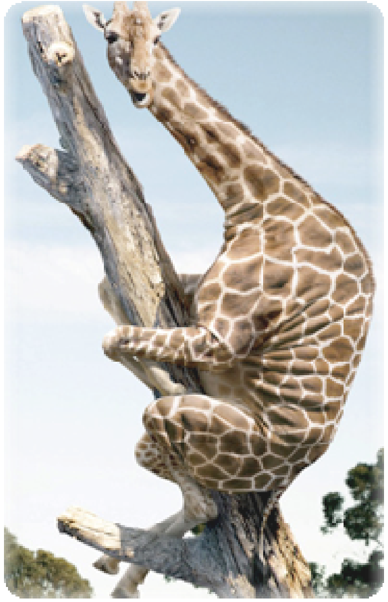
Hierarchical spatial-temporal feature learning

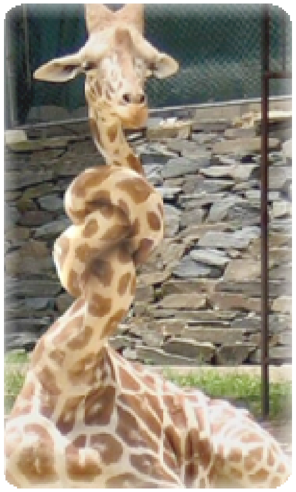
Observed gaze sequence

Model predictions

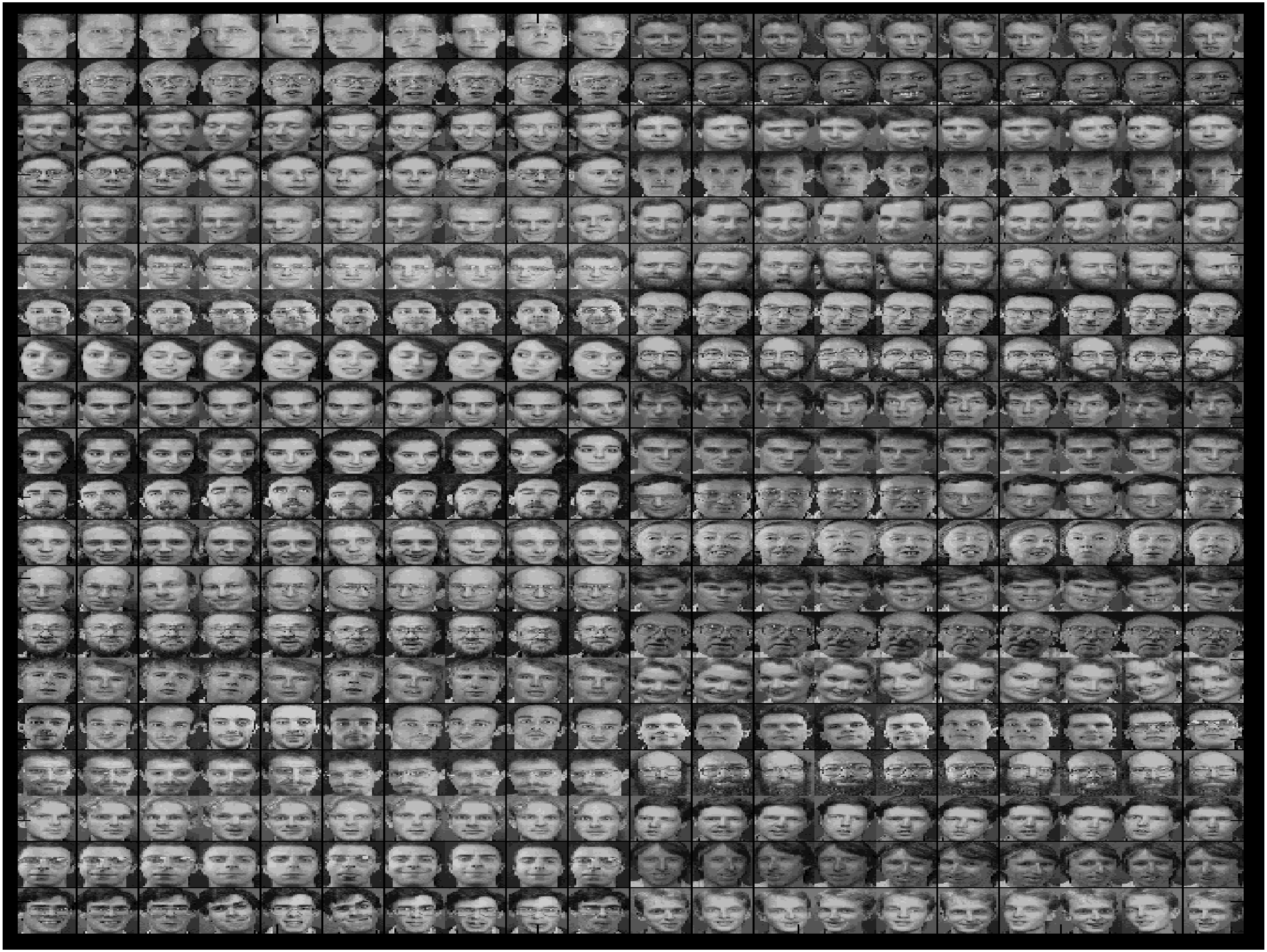
[Bo Chen et al 2010]

Application: Invariant recognition in natural images

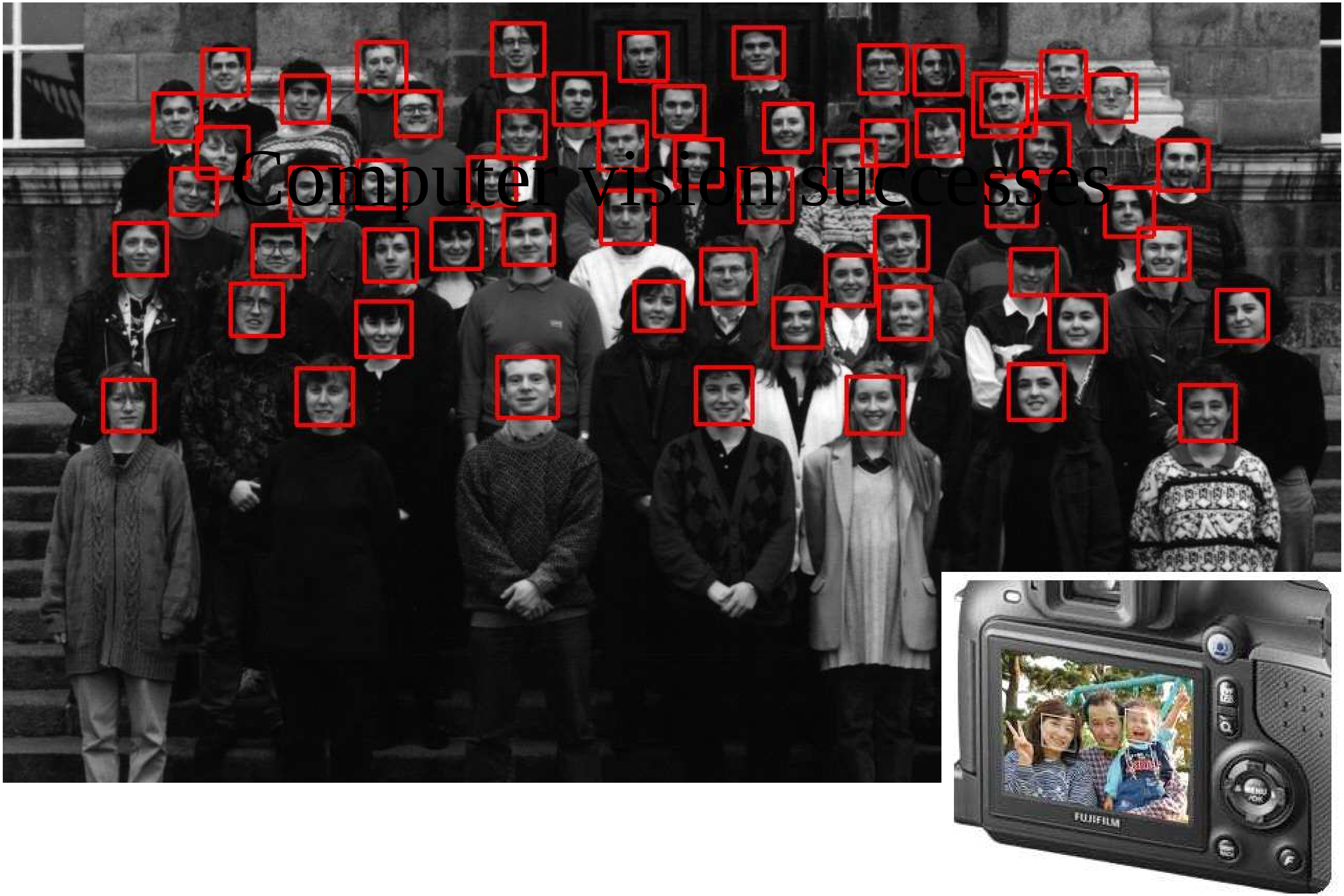




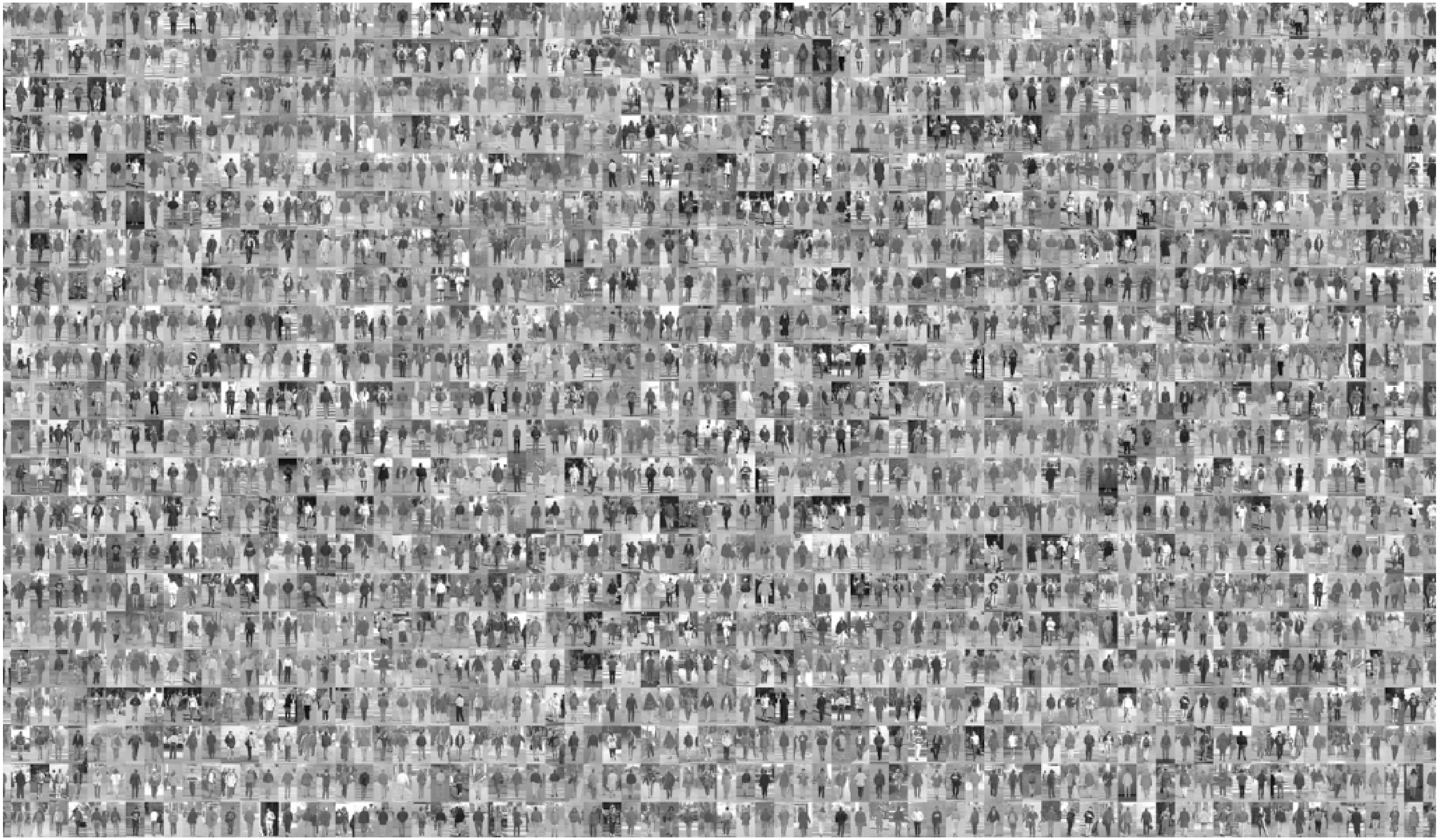
[Thomas Serre 2012]



Computer vision successes



[Thomas Serre 2012]



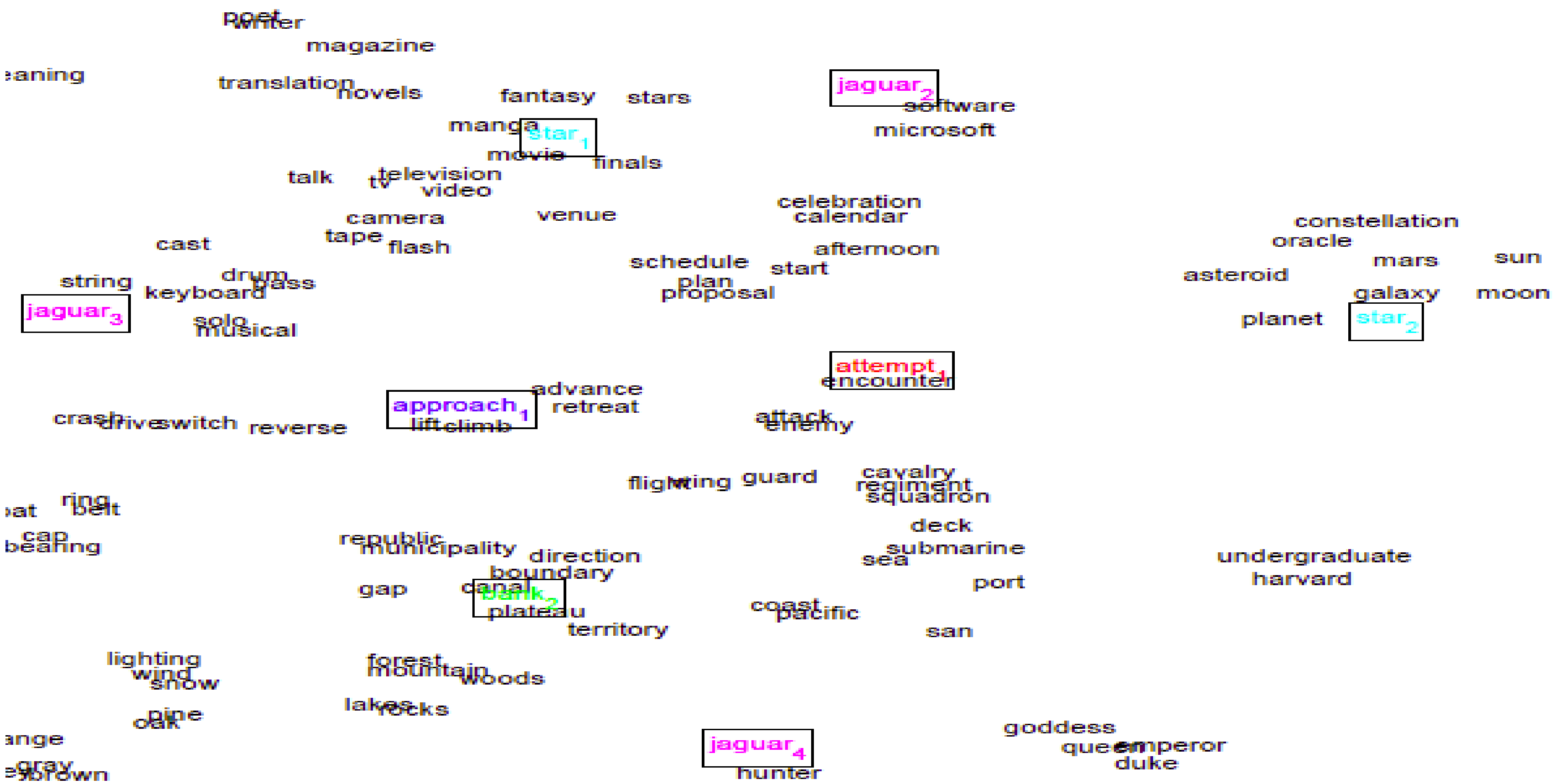
Millions of labeled examples are used to build real-world applications, such as pedestrian detection

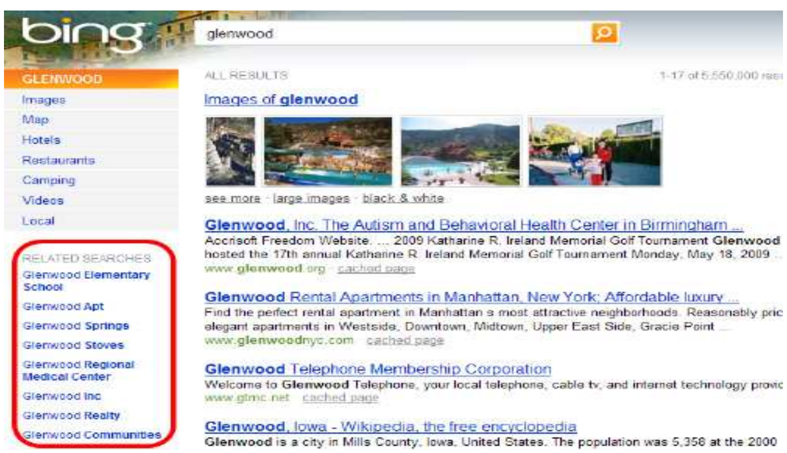
[Tomas Serre]

Application: Autonomous driving

Mobileye: Already available on Volvo S60 and soon on most car manufacturers

Application: Information Extraction

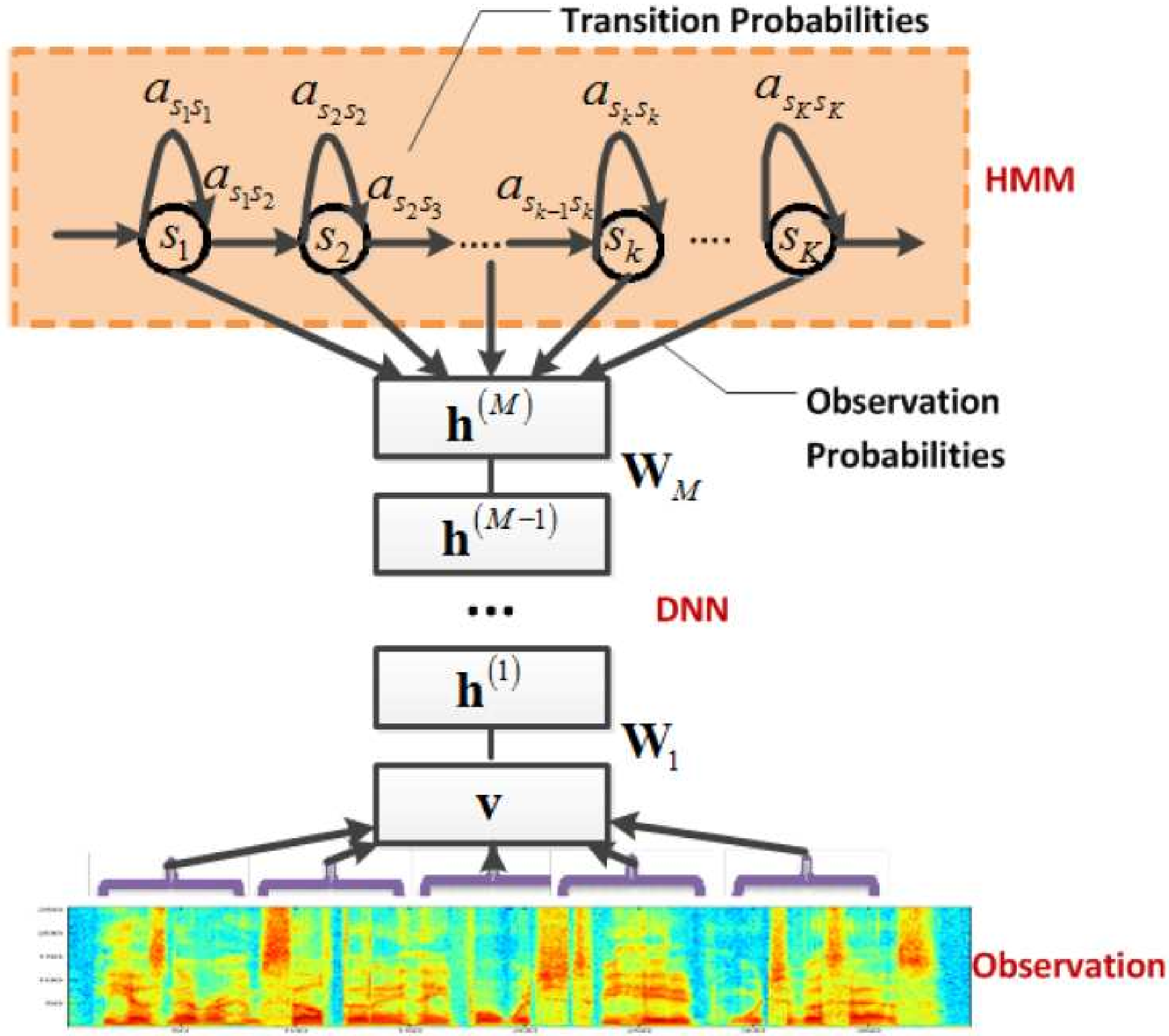




e^l	people				
r	build	destroy	won	suffer	control
e^r	*livelihoods*	*icons*	*emmy*	*sores*	*rocket*
	homes	*virtue*	*award*	*agitation*	*stores*
	altars	*donkeys*	*everything*	*treatise*	*emotions*
	houses	*cowboy*	*standings*	*eczema*	*spending*
	ramps	*chimpanzees*	*pounds*	*copd*	*fertility*

[Yoshua Bengio, Jason Weston, Richard Socher]

Application: Speech recognition



[George Dahl et al 2011]

Next lecture

In the following lecture we will begin to learn the probabilistic tools we need to ***understand*** machine learning and ***innovate*** algorithms, models and applications.