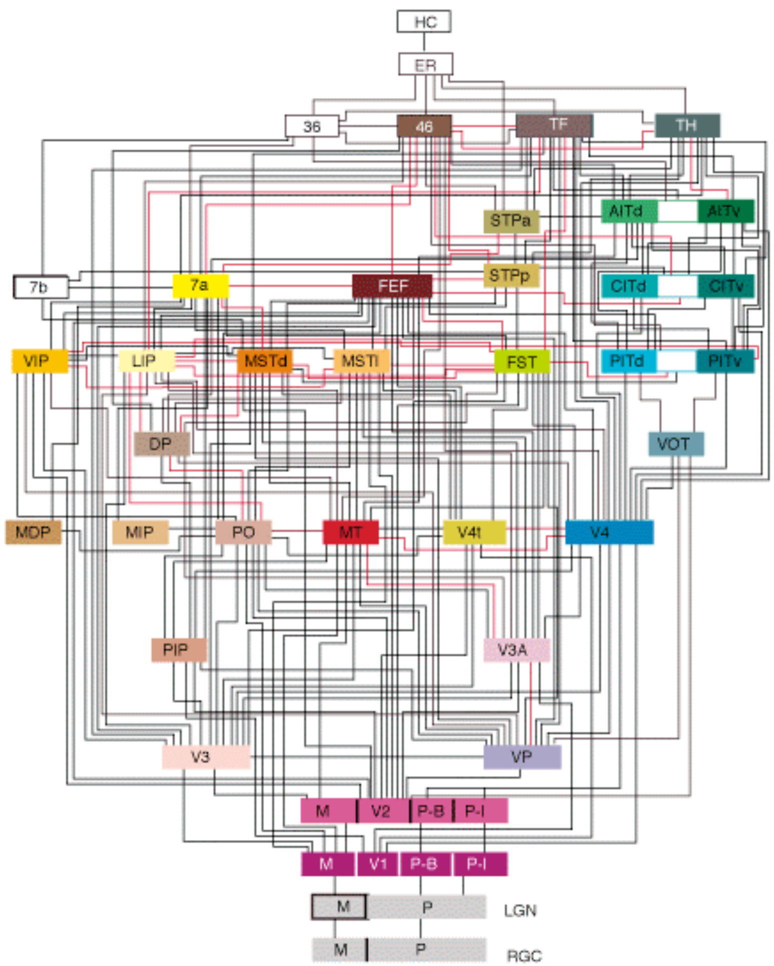
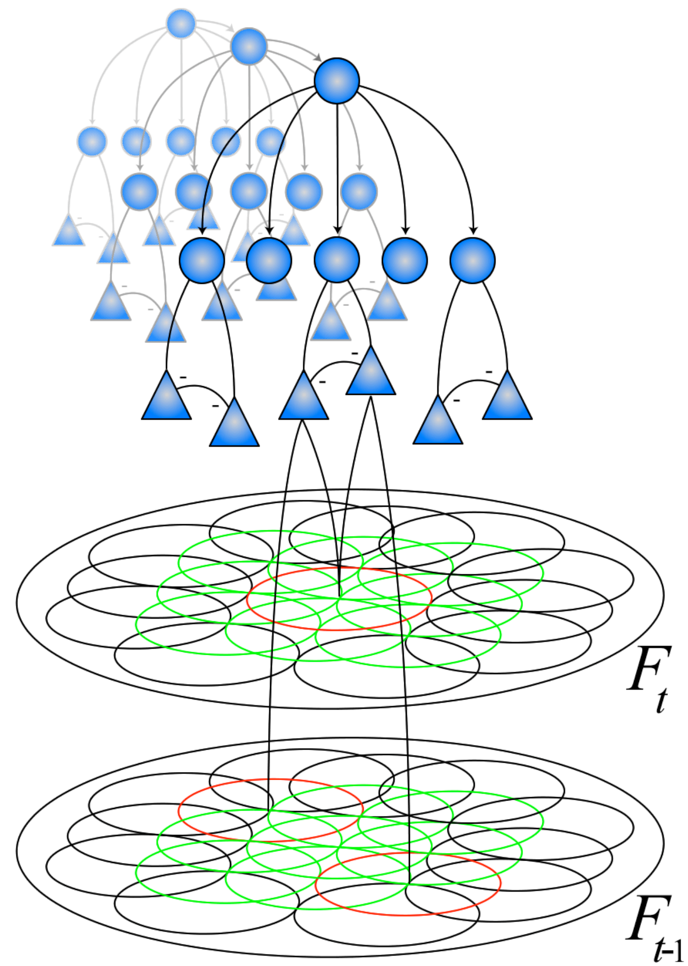
Computational Models of the Primate Visual Cortex

Thomas Dean
Google

Hierarchical Graphical Models

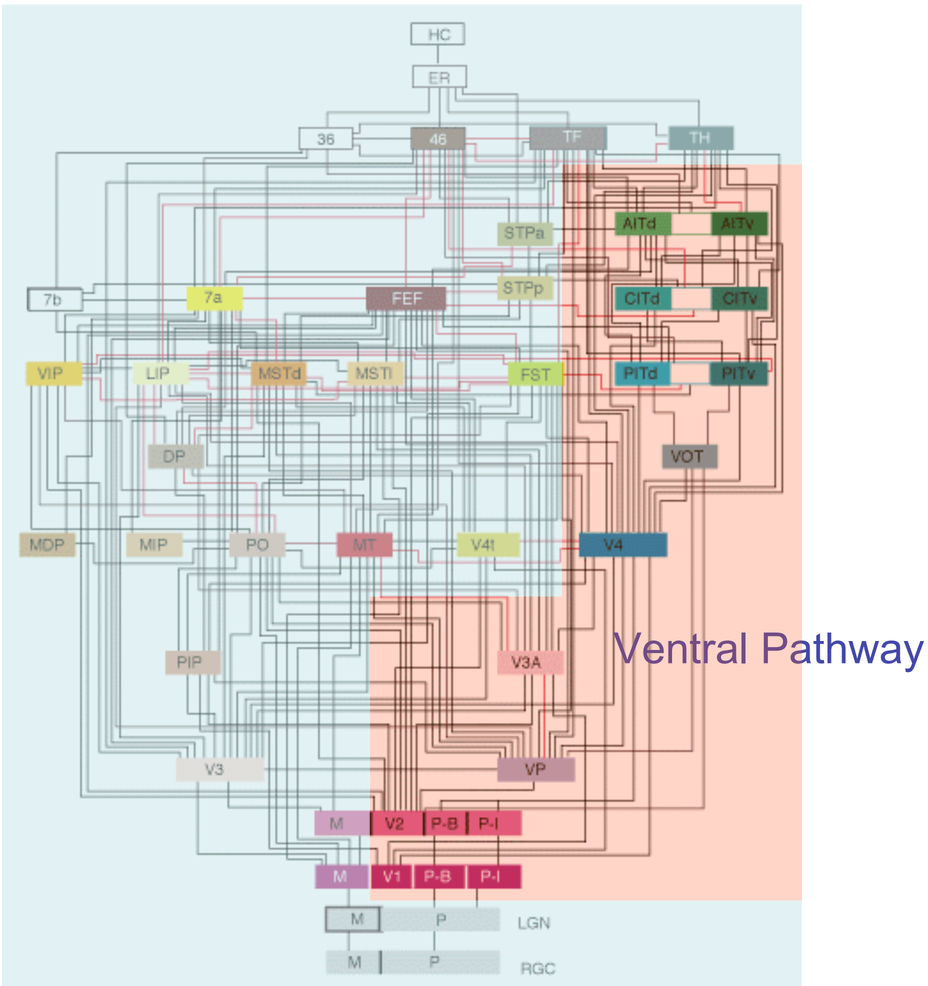


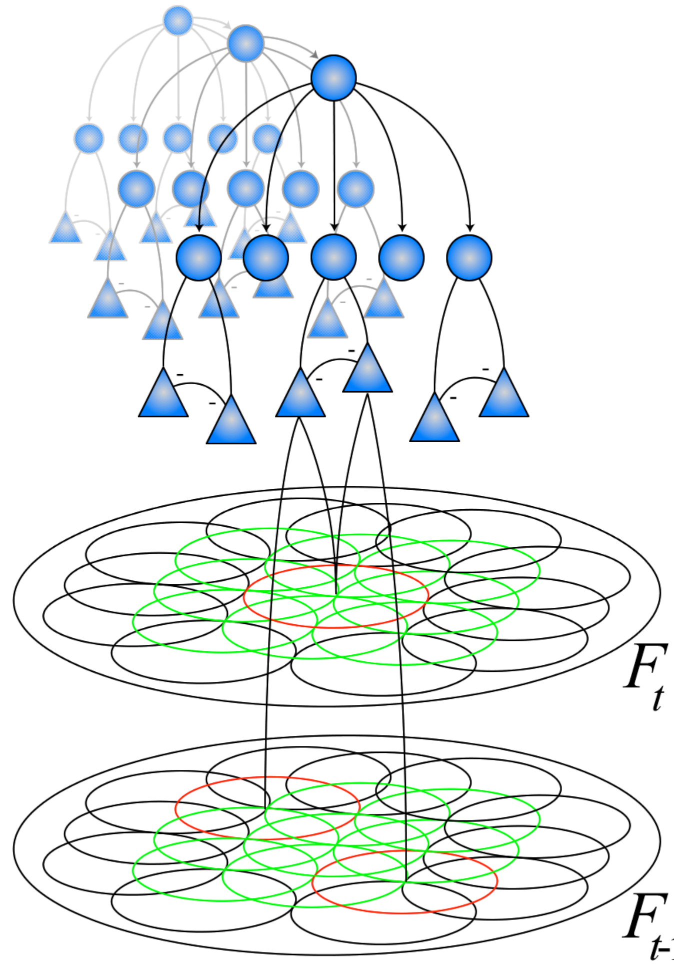


Felleman, D. J. and Van Essen, D. C. *Distributed hierarchical processing in primate cerebral cortex*. Cerebral Cortex, 1:1-47, 1991.

Thomas Dean. *Learning invariant features using inertial priors*. Annals of Mathematics and Artificial Intelligence, 47(3-4):223-250, 2006.

Hierarchical Graphical Models

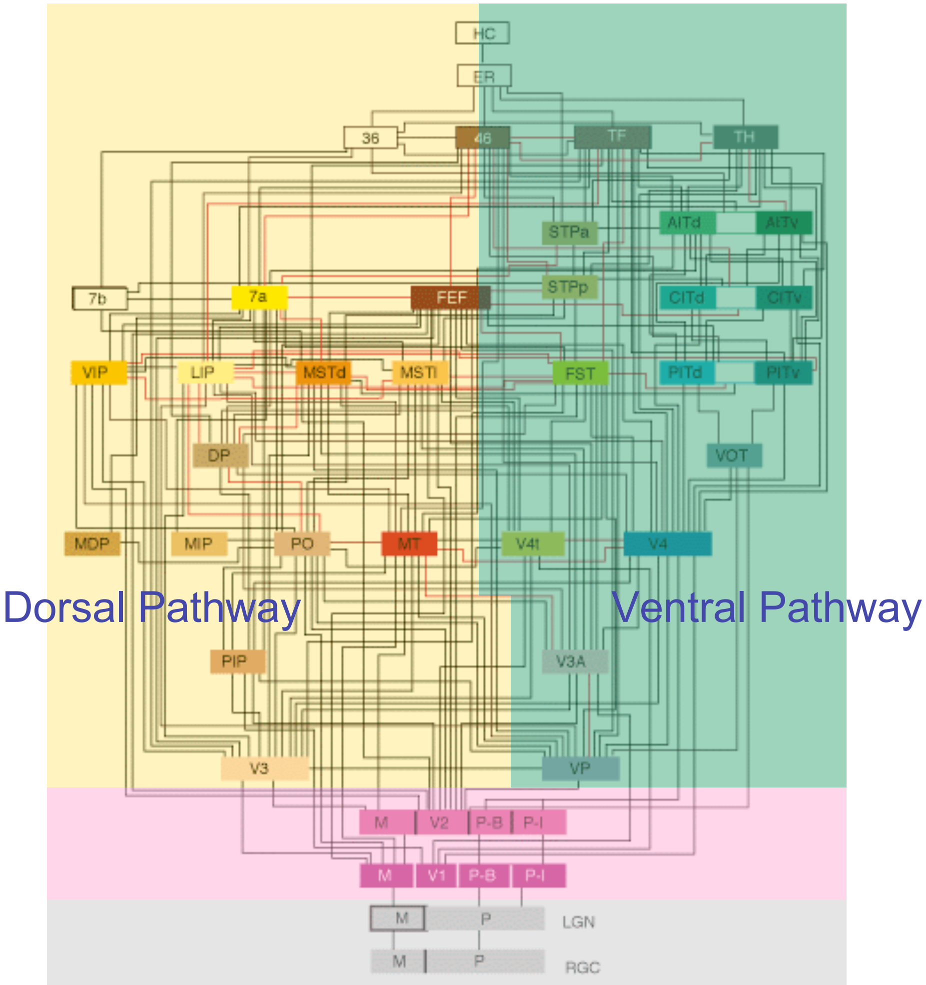


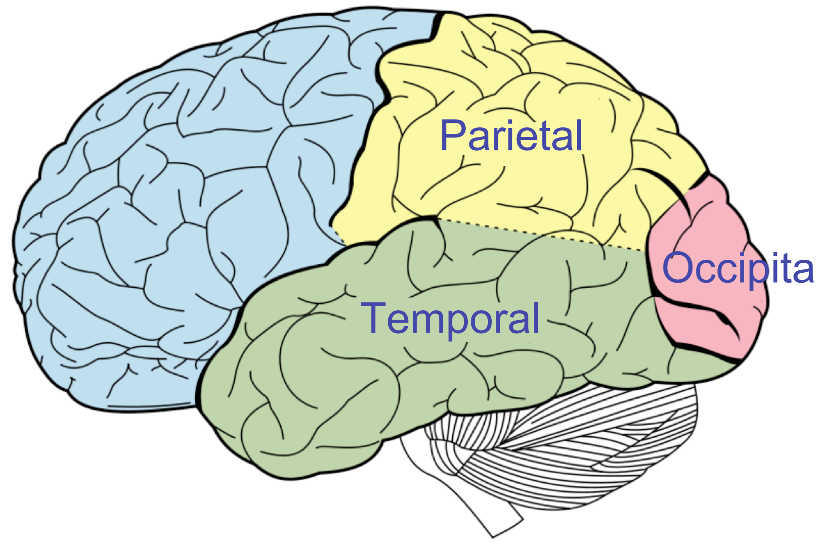


Felleman, D. J. and Van Essen, D. C. *Distributed hierarchical processing in primate cerebral cortex*. Cerebral Cortex, 1:1-47, 1991.

Thomas Dean. *Learning invariant features using inertial priors*. Annals of Mathematics and Artificial Intelligence, 47(3-4):223-250, 2006.

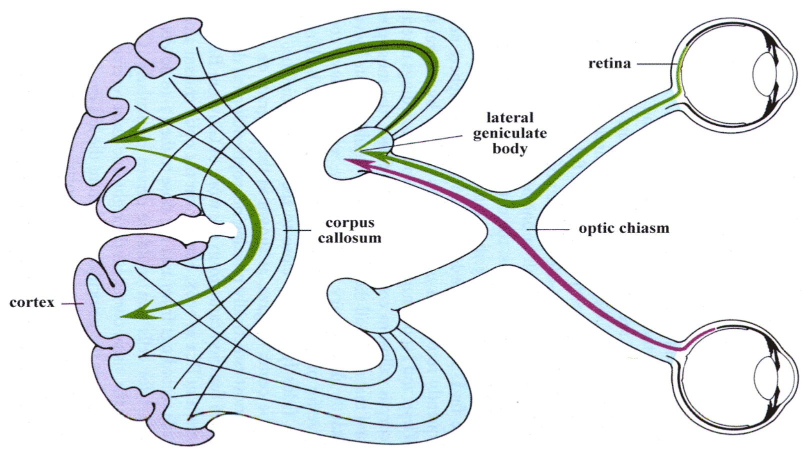
Hierarchical Graphical Models

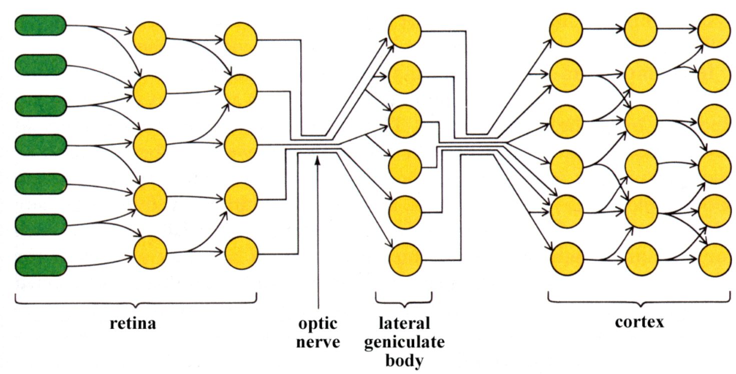


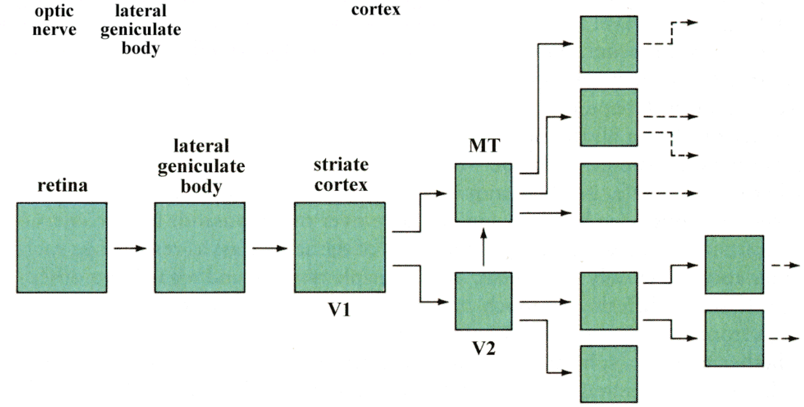


Felleman, D. J. and Van Essen, D. C. *Distributed hierarchical processing in primate cerebral cortex*. Cerebral Cortex, 1:1-47, 1991.

Terminology: Receptive Fields

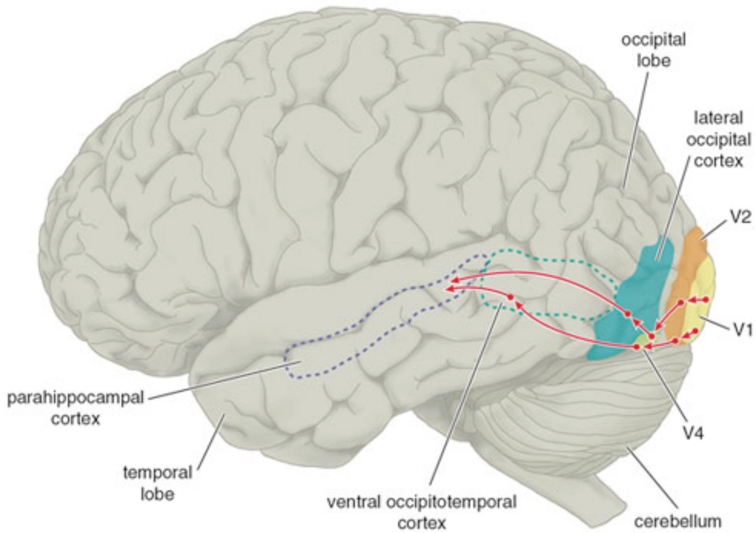




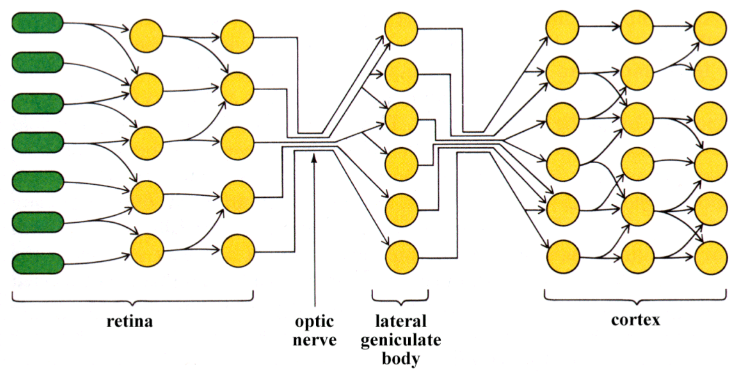


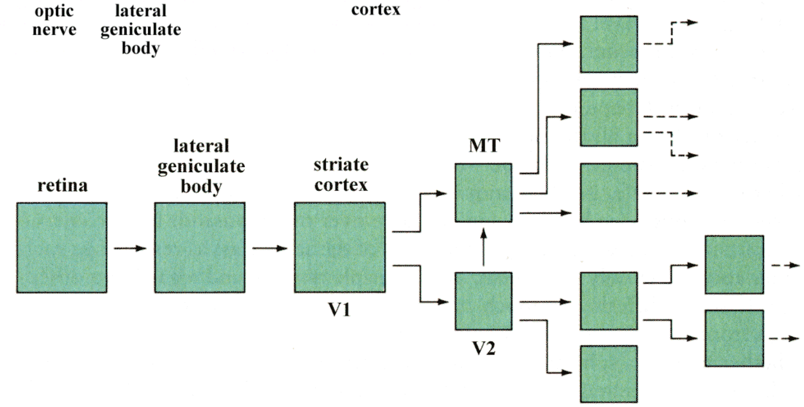
† [Hubel and Wiesel, 1962]

Terminology: Retinotopic Maps



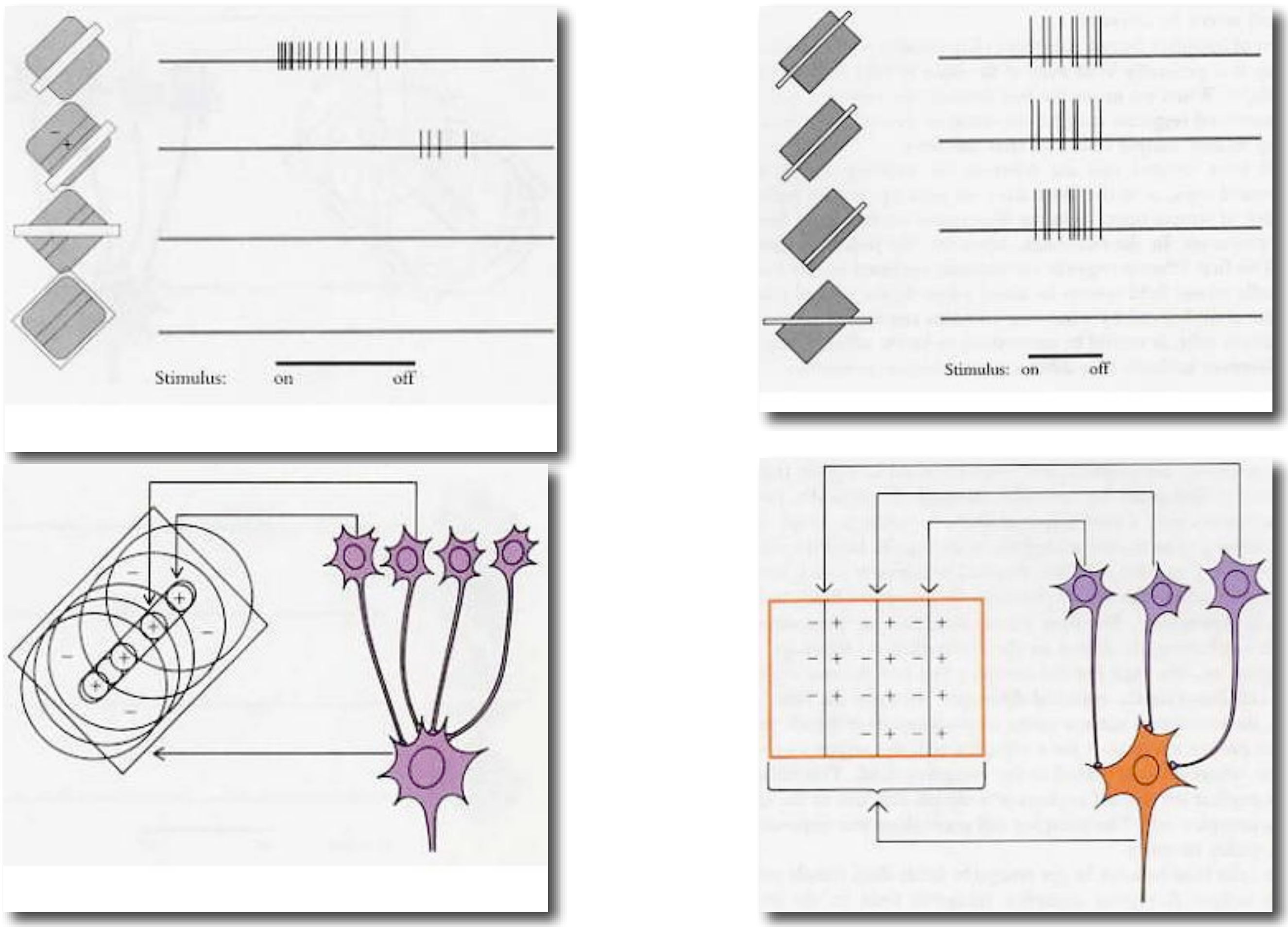
ventral visual pathway





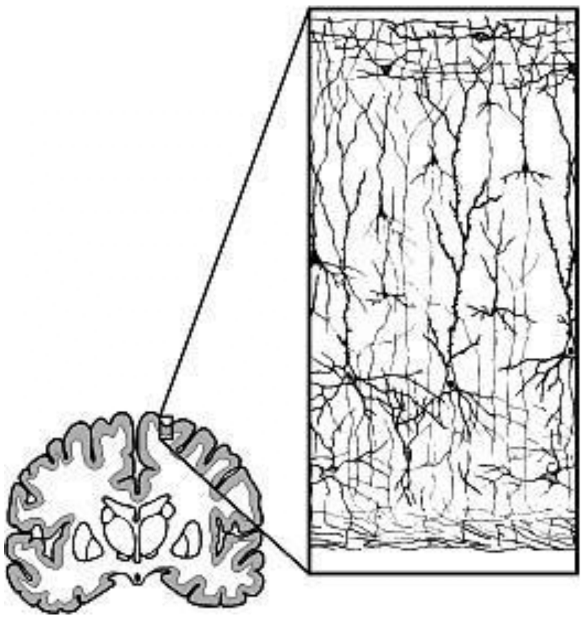
† [Hubel and Wiesel, 1962]

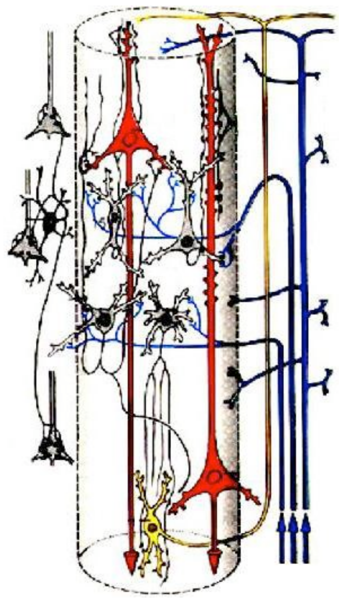
Simple and Complex Cells

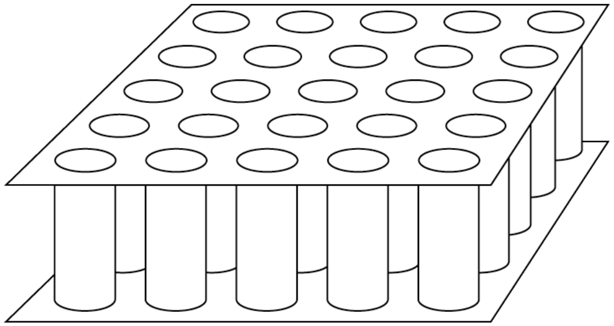


† [Hubel, 1988]

Terminology: Cortical Columns

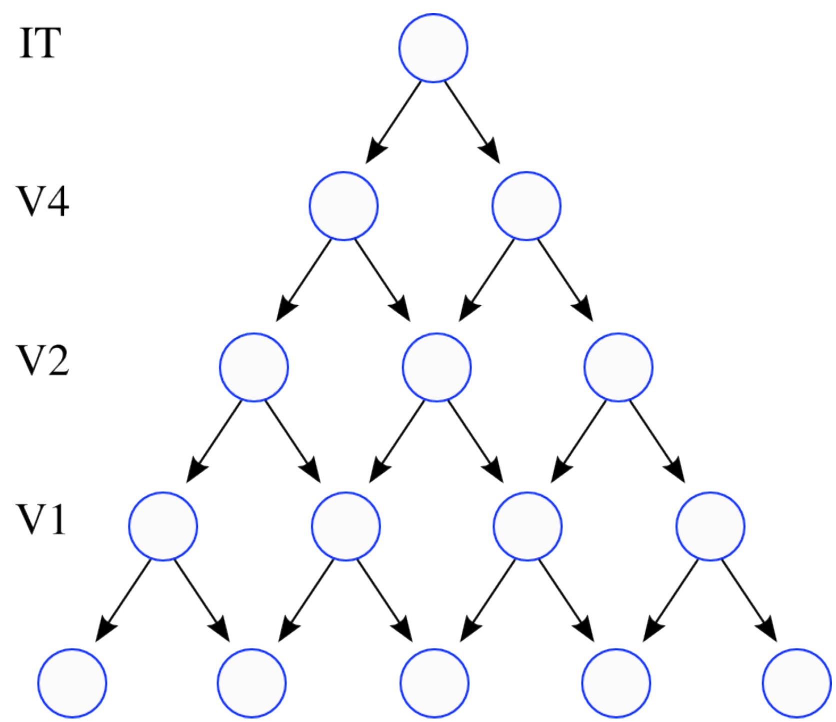






† [Mountcastle, 1957]

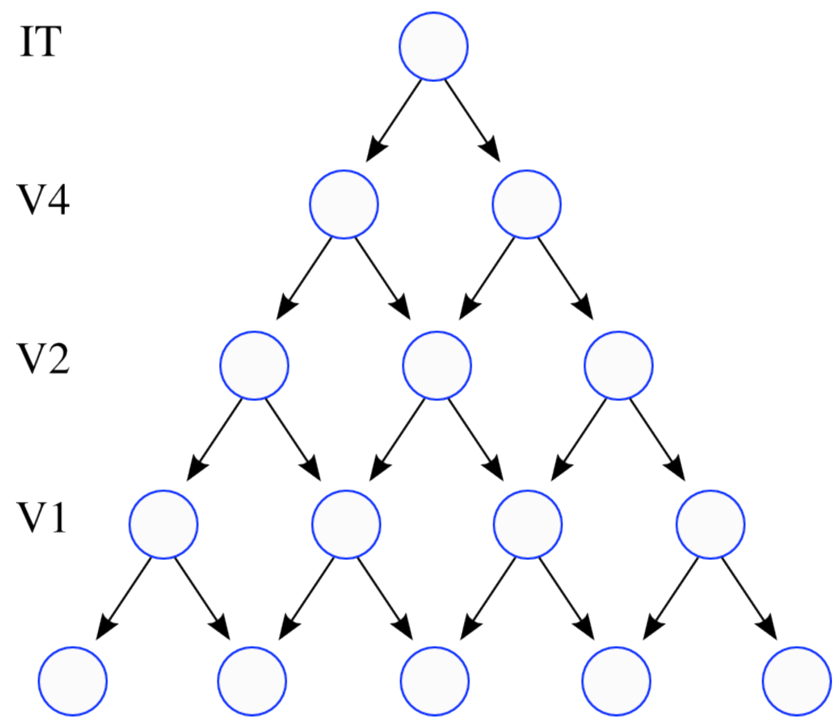
Simple Hierarchical Cortex Model[1]



1. Ignores intra-layer edges, edges that span layers, explicit feedback, locality, and organization reflecting spatial relationships in peripheral nervous system.

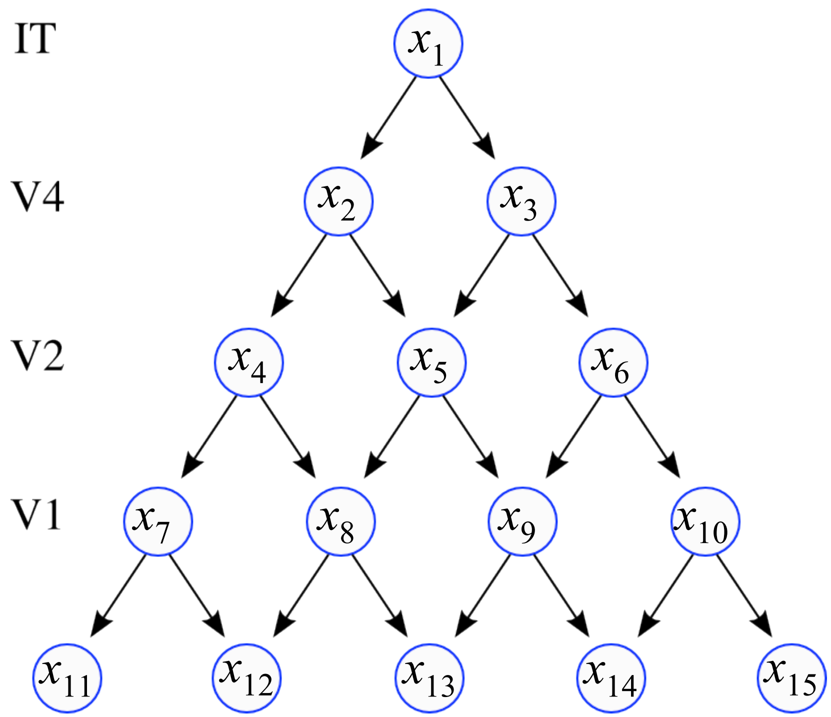
Simple Hierarchical Cortex Model[1]

Specify a predefined[1] network structure capable of encoding a simple hierarchy



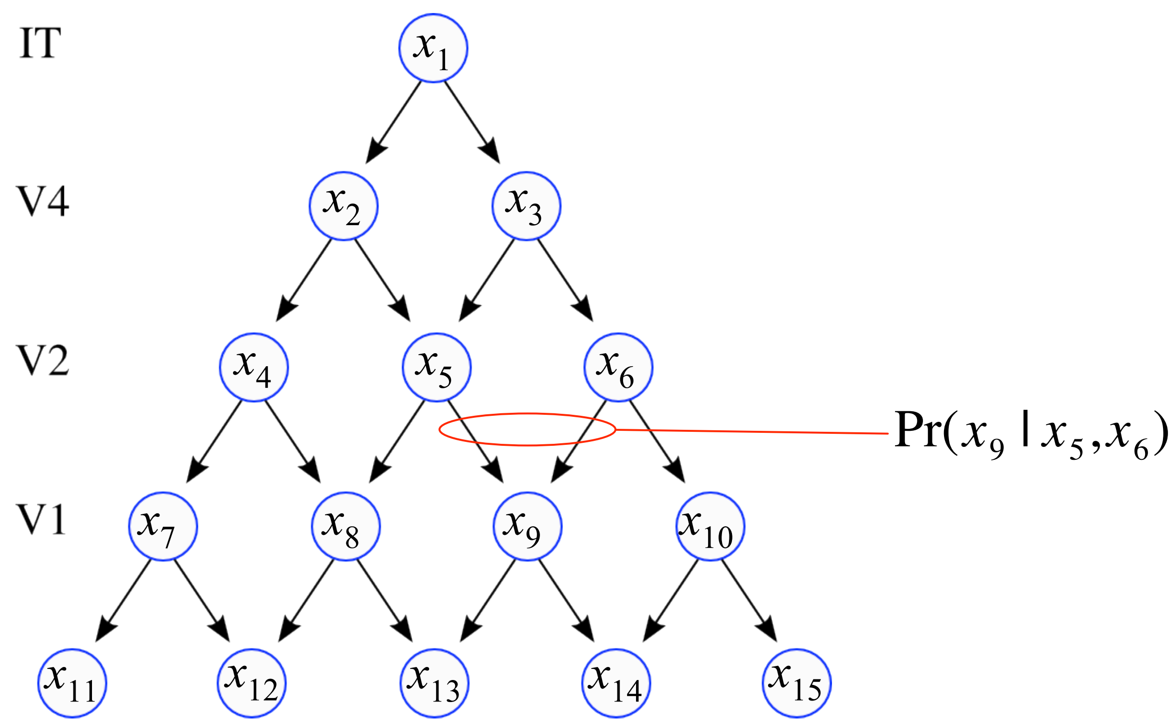
1. One approach is to over determine the network structure by generating a *small-world* graph and using parameter learning to quantify connections.

Probabilistic Graphical Models[1]



1. Graph in which nodes correspond to random variables and edges correspond to dependency relations quantified by specifying conditional probability functions.

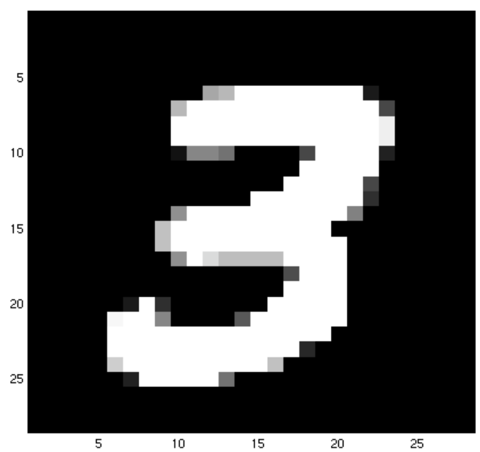
Probabilistic Graphical Models[1]

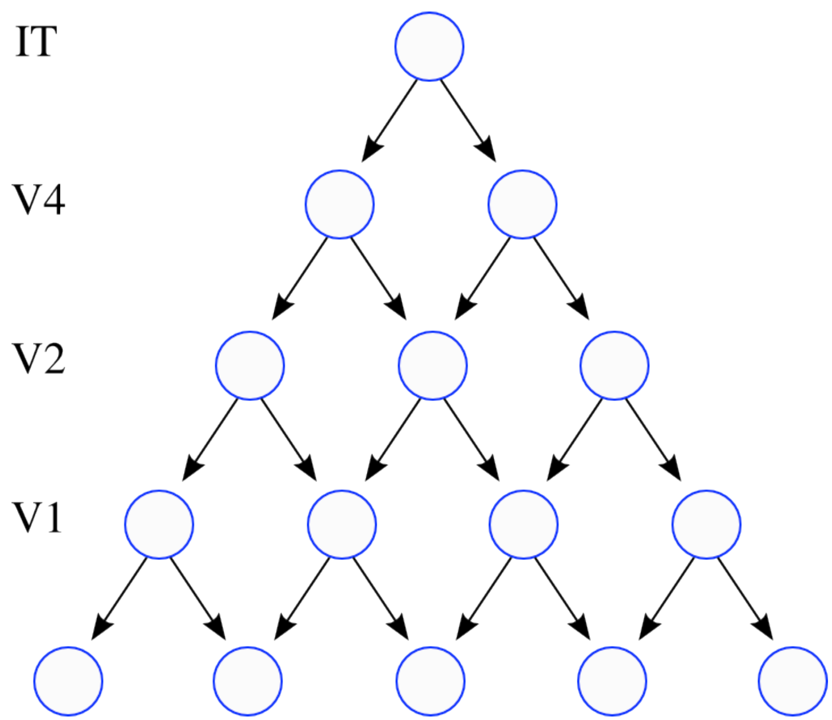


1. Compactly describes the full joint probability: $\Pr(x_1,\ldots,x_n) = \prod_{i=1}^{n} \Pr(x_i \mid \text{Parents}(x_i))$

Inference in Graphical Models

- Learn to recognize hand-written digits



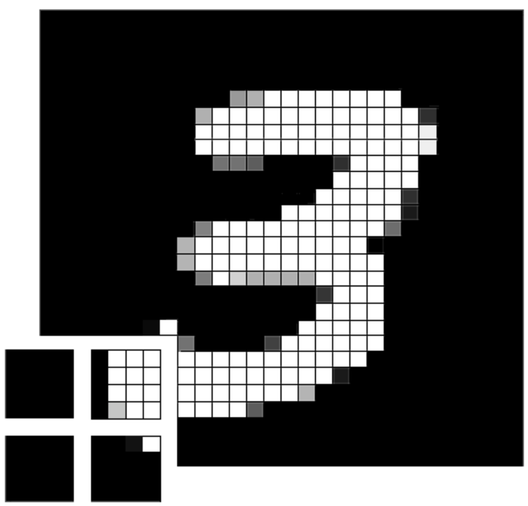


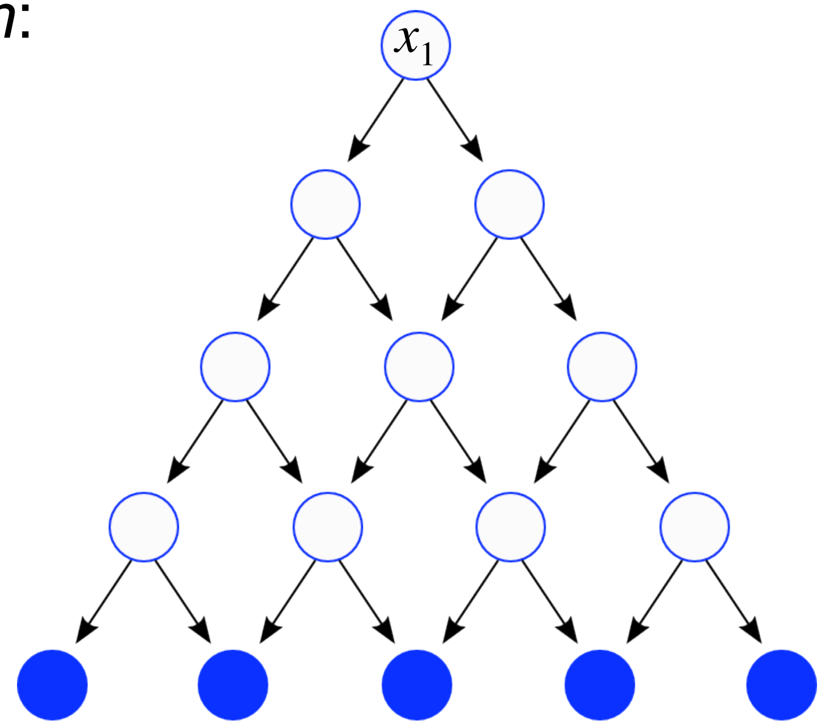
Inference in Graphical Models

- Observe input sensor data S
- Compute[1] the *belief function*:

$$\forall i,\ \mathrm{Bel}(x_i) = P(x_i \mid S)$$

$$\Omega_{x_1} = \{0,1,2,\ldots,9\}$$



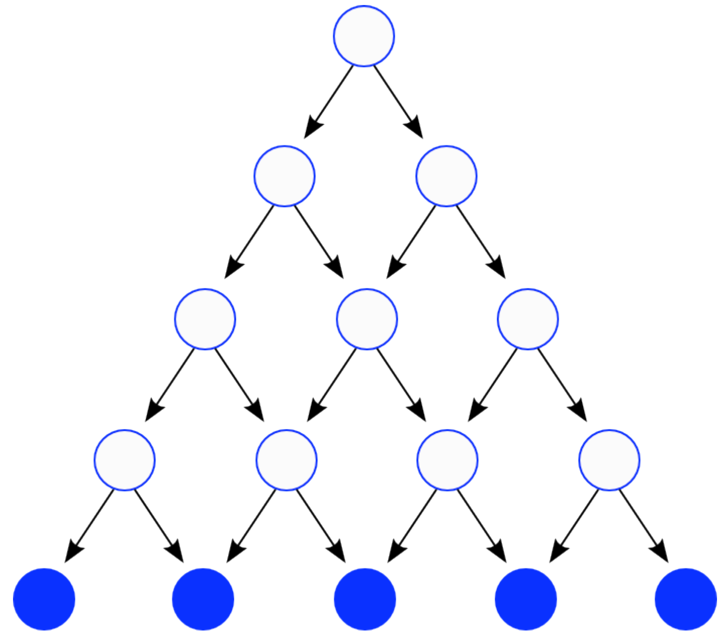


1. Loopy belief propagation and nonparametric belief propagation / particle filtering are ideally suited for iteratively updating this belief function.

Learning in Graphical Models

- Mixtures of Gaussians[1]

$$\Pr(\mathbf{x} \mid \theta = i) = \sum_{j=1}^{K} \alpha_{ij} \mathrm{N}(\mathbf{x} \mid \mu_k, \Sigma_k)$$



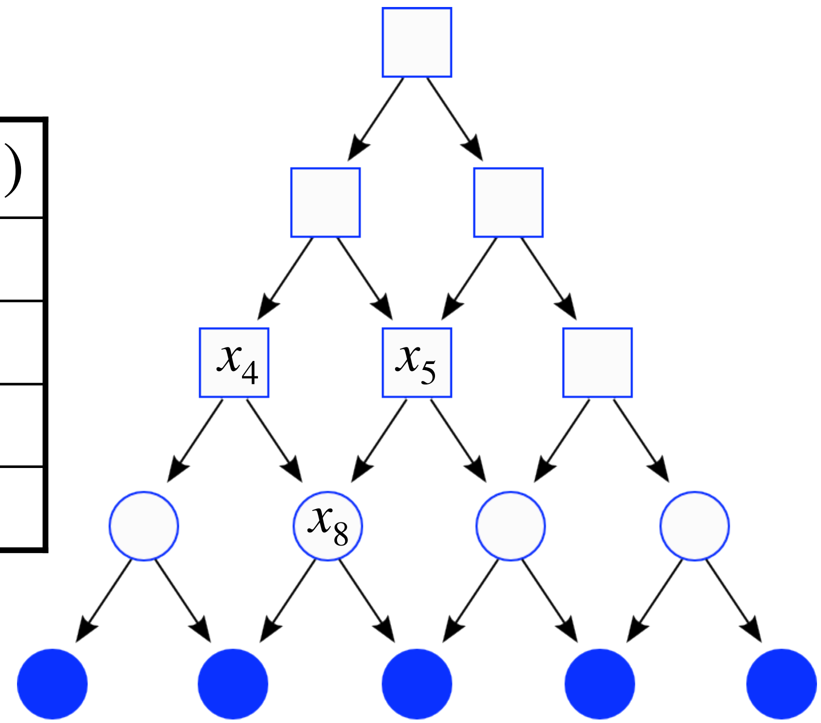
1. There are many possible choices to model real-valued inputs. Gaussian mixtures are simple to learn and provide a discrete conditioning variable.

Learning in Graphical Models

- Learn the remaining layers using bottom-up layer-by-layer method[1]

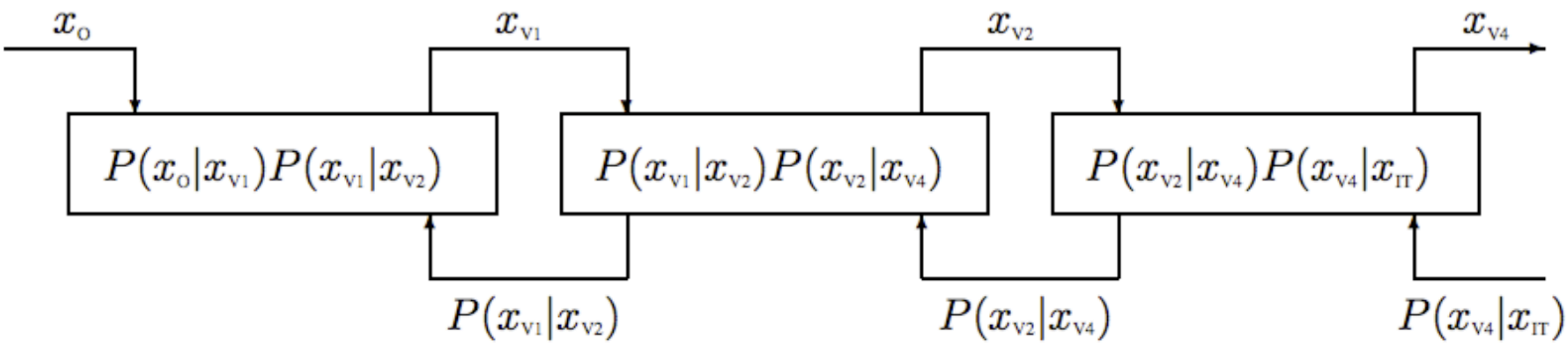
Ω_{x_i} , $\Pr(x_i \mid \text{Parents}(x_i))$

x_4	x_5	x_8	$P(x_8 \mid x_4, x_5)$
1	1	1	0.147
2	1	1	0.053
3	1	1	0.238
…		…	…



1. This layer-by-layer strategy has a long history in artificial neural networks and machine learning and currently a focus of interest in *deep belief networks*.

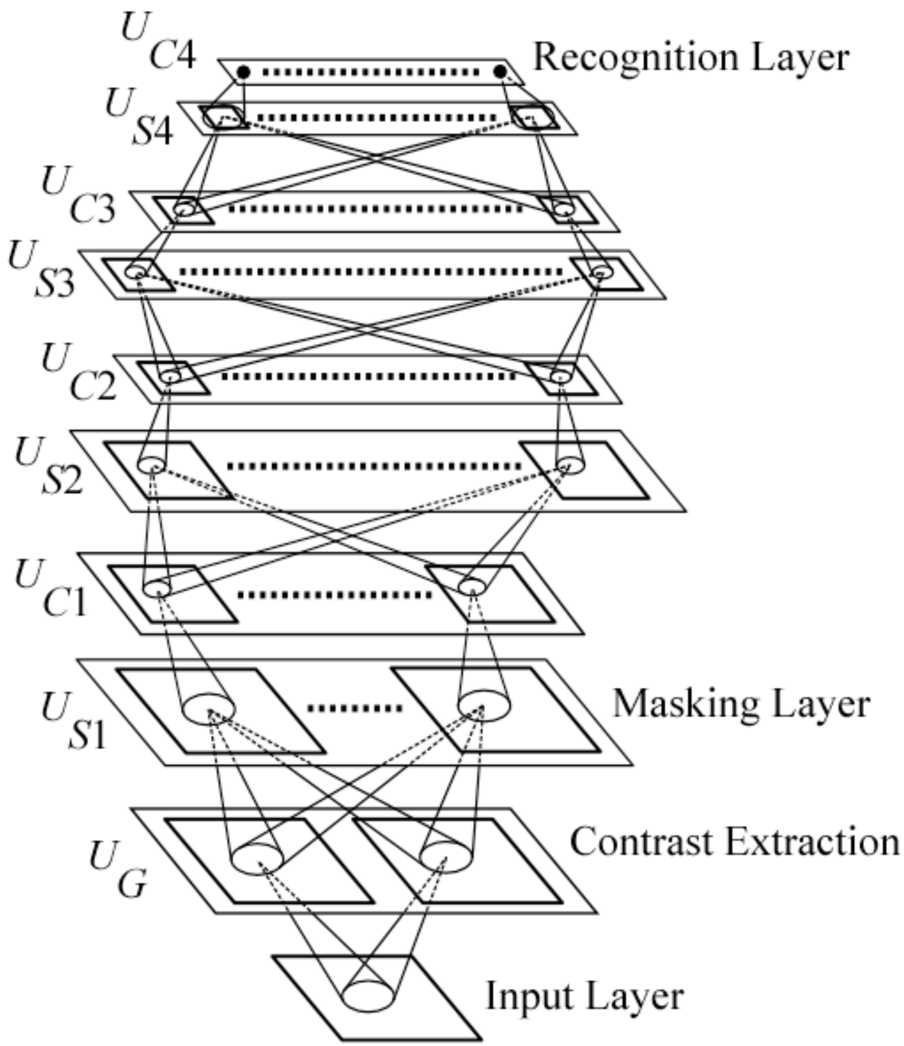
Lee and Mumford Model



- Hierarchical model of the visual cortex
- Generative Bayesian statistical model
- Bottom-up data $\{x_i\}$ are fed forward
- Top-down priors $\{P(x_i|x_{i+1})\}$ are fed back

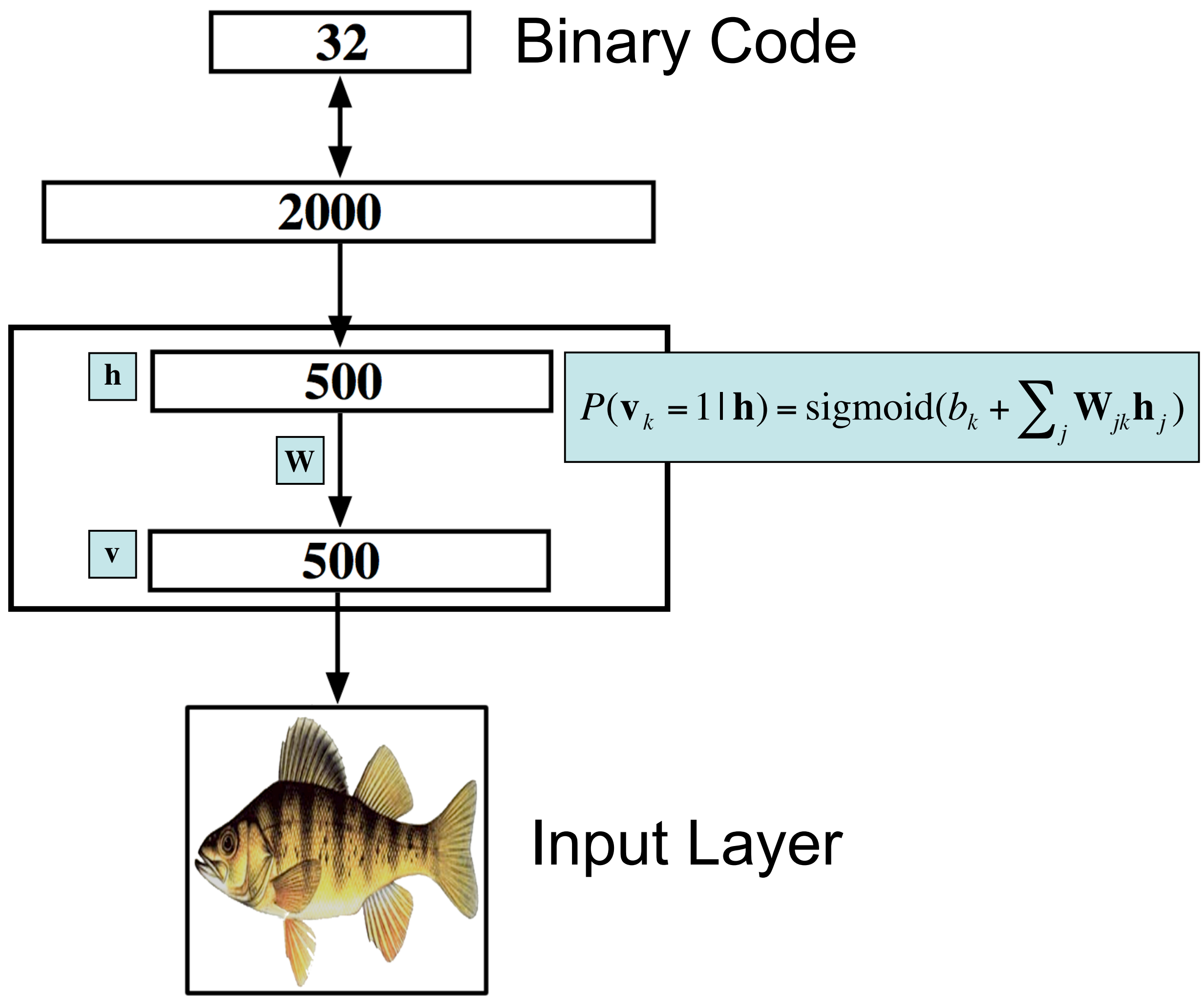
Lee, T. S. and D. Mumford (2003). Hierarchical Bayesian inference in the visual cortex. Journal of the Optical Society of America 2 (7), 1434-1448.

Multilayer Perceptron Models



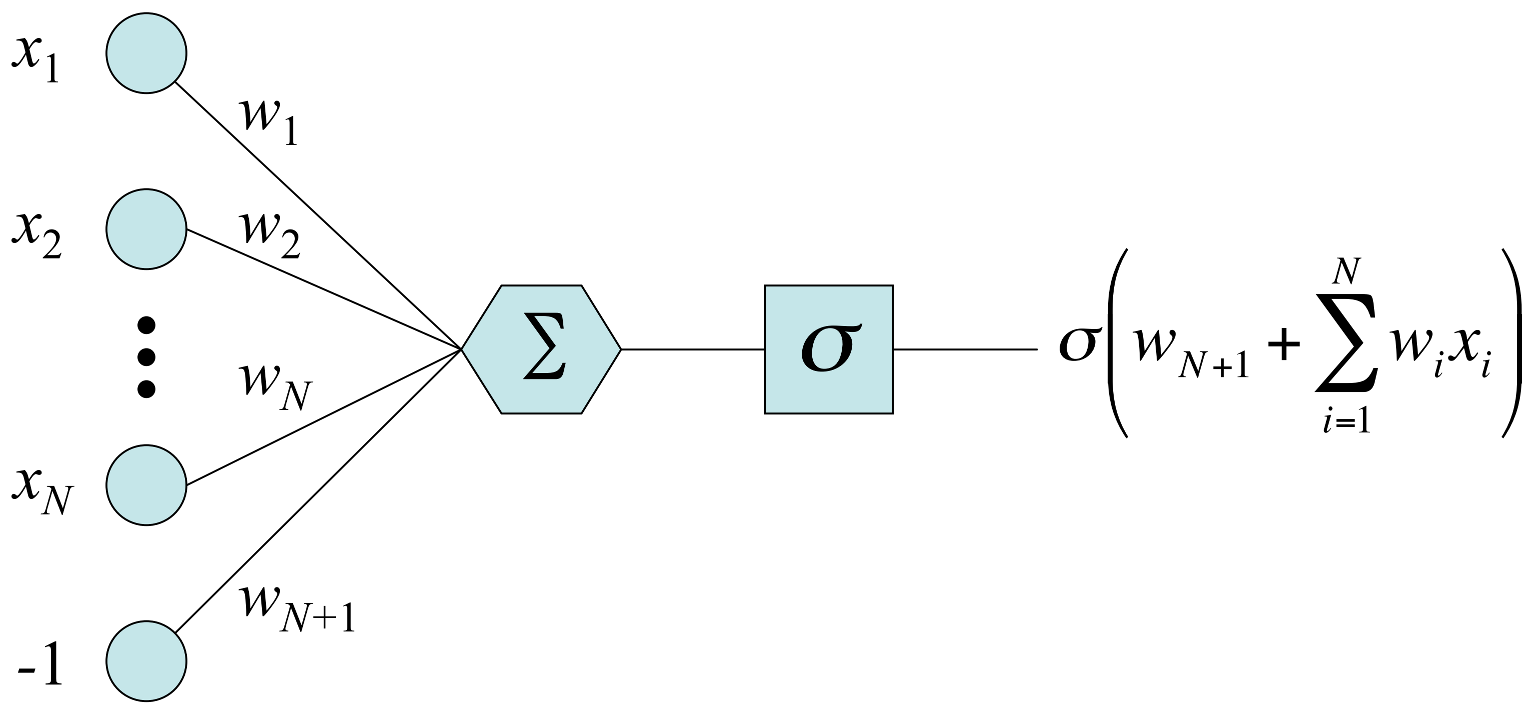
K. Fukushima. Neocognitron: A self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics, 36(4):93–202, 1980.

Deep Belief Networks



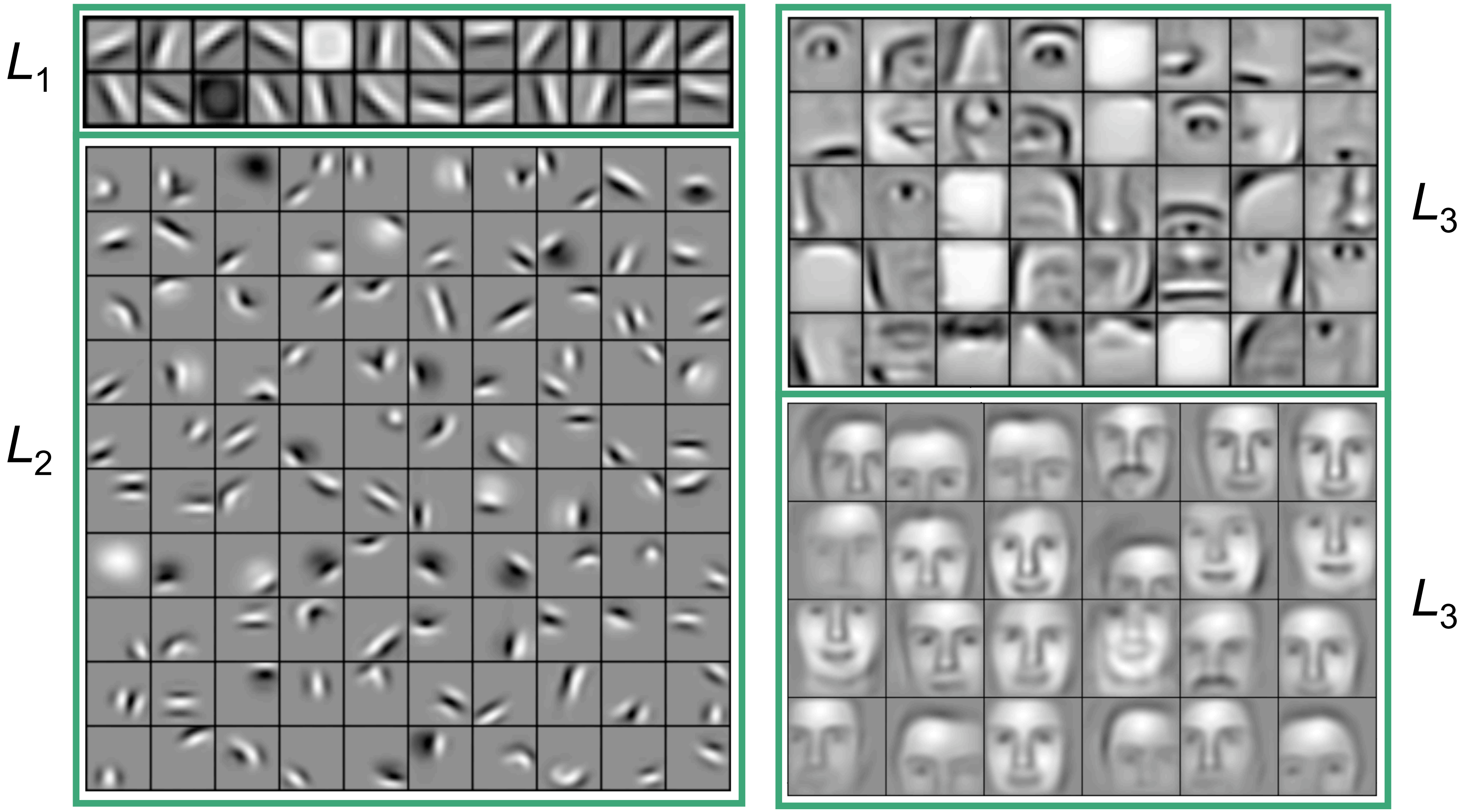
Hinton, Geoffrey. E. (2005). What kind of a graphical model is the brain? In the Proceedings of the 19th International Joint Conference on Artificial Intelligence.

McCulloch and Pitts Neuron

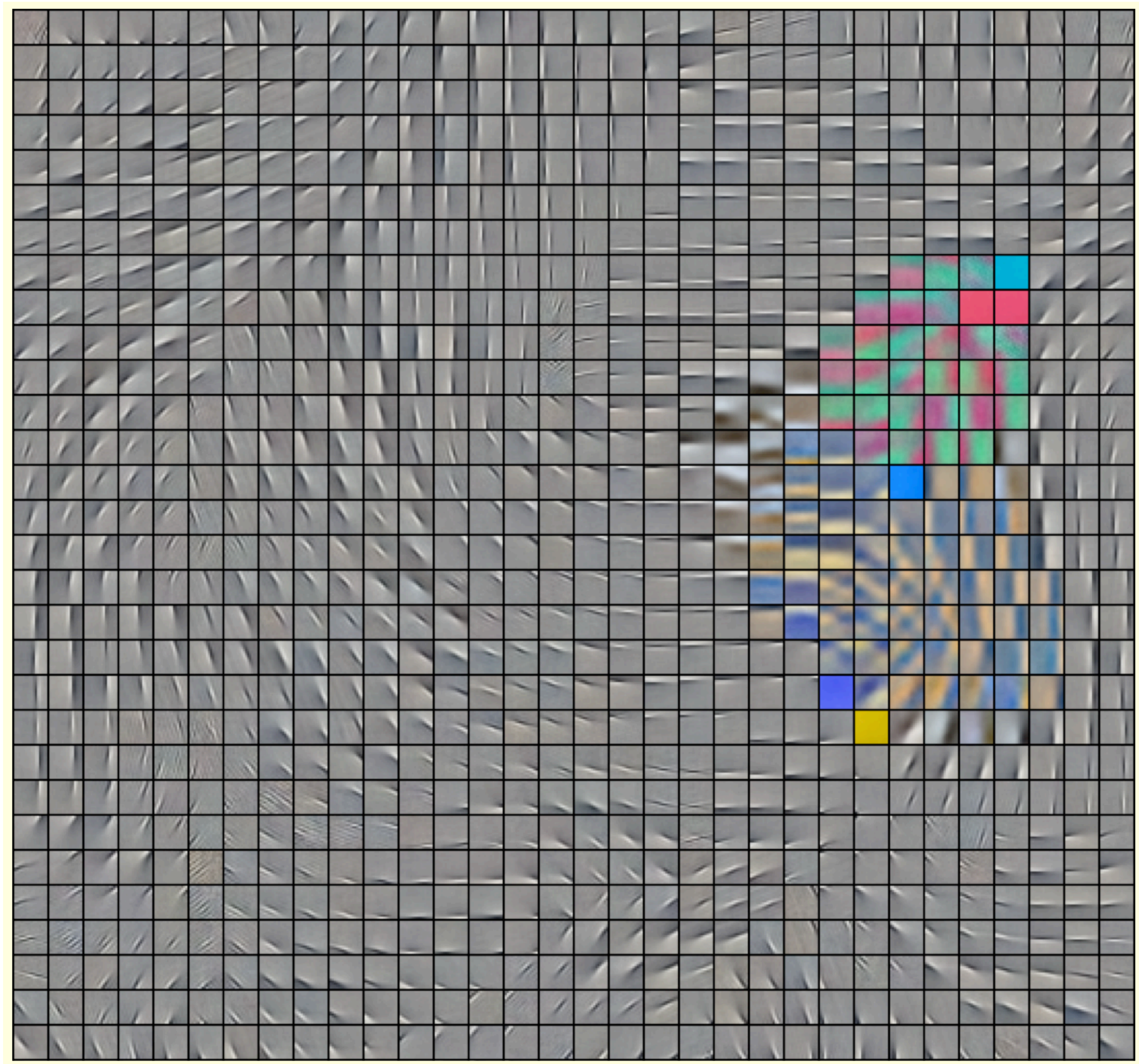


McCulloch, W. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*. 7:115-133.

Modeling Interesting Phenomena



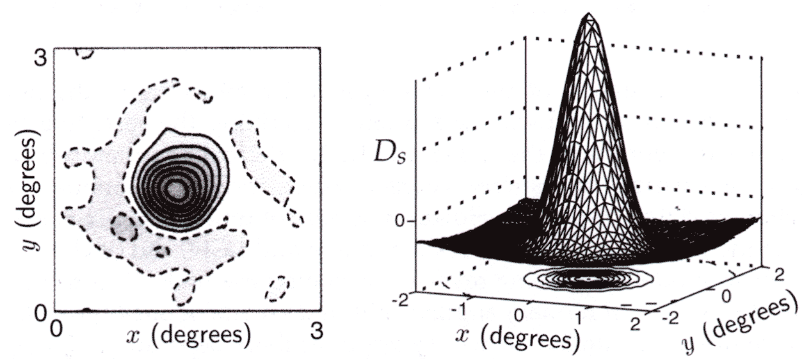
Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. Proceedings of the 26th Annual International Conference on Machine Learning, 2009.



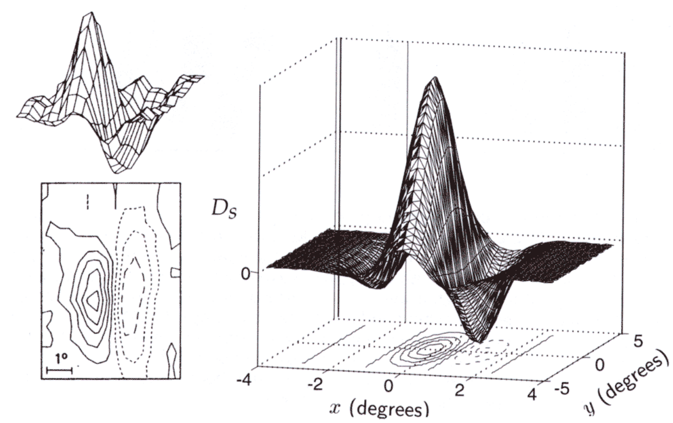
Marc' Aurelio Ranzato and Geoffrey Hinton. Deep learning with multiplicative interactions. In the Proceedings of the *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2010.

Simple Cell Receptive Fields

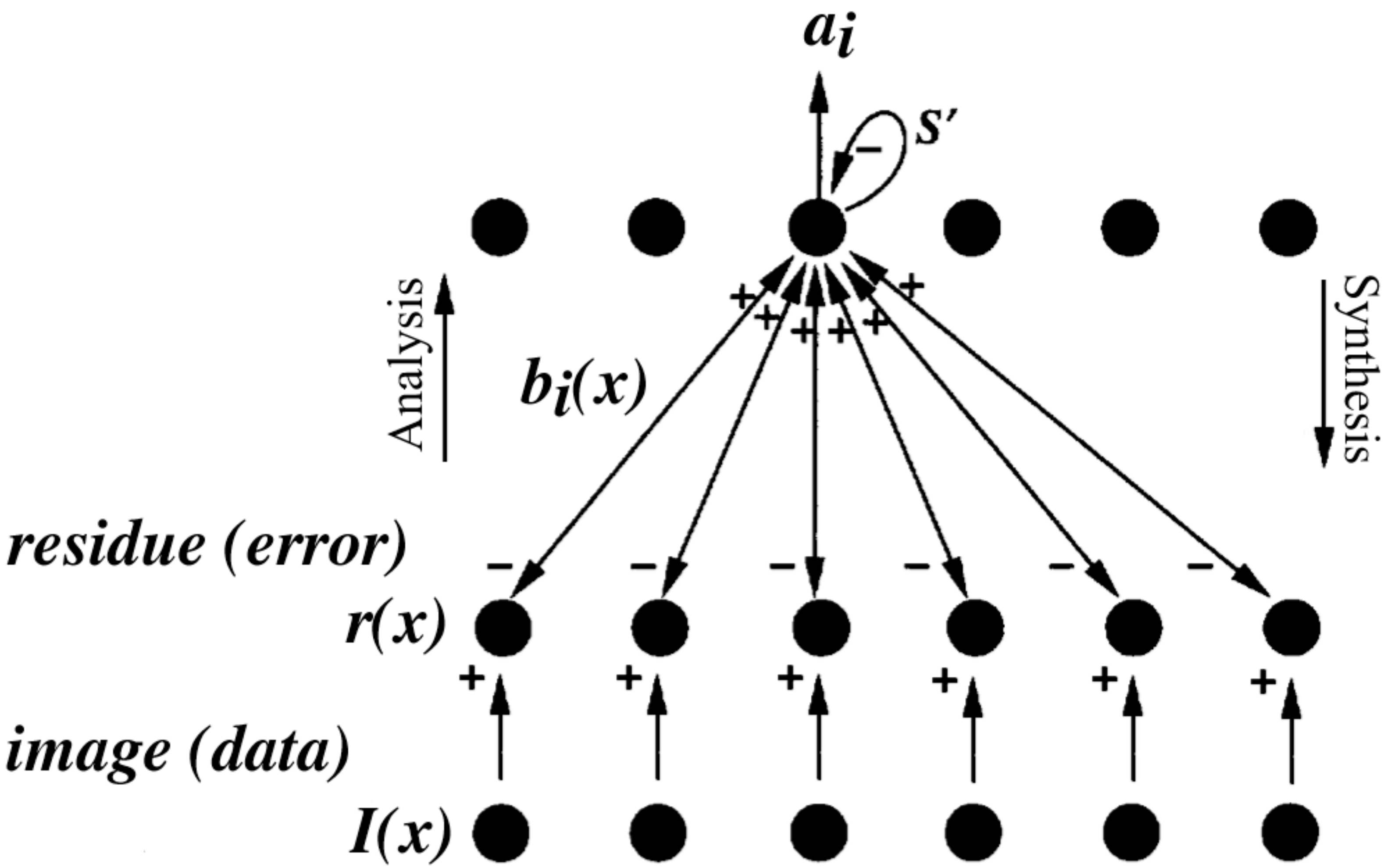
- Center-surround cells in the retina and lateral geniculate nuclei — difference of Gaussians



- Edge-sensitive cells in V1 — product of Gaussian and sinusoidal functions called Gabor functions

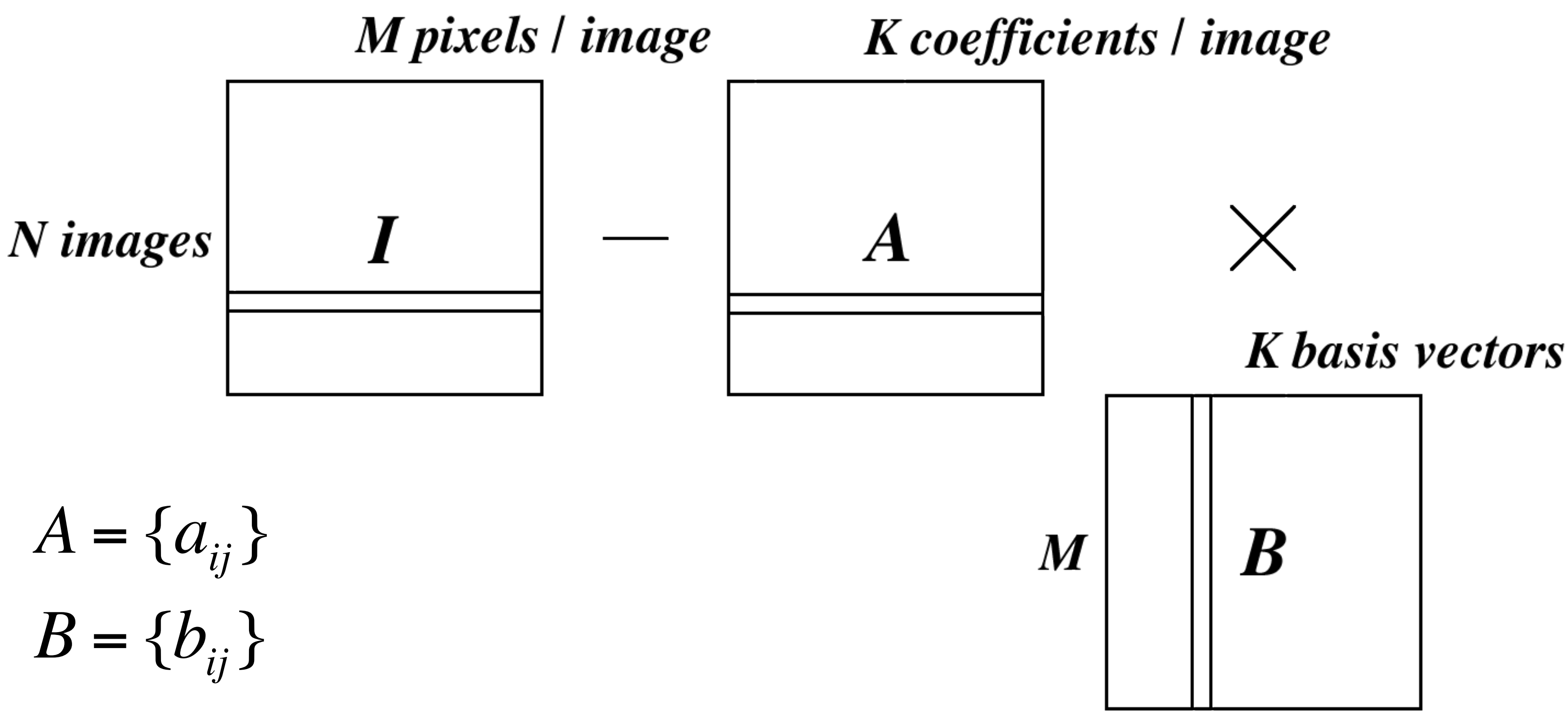


Analysis Synthesis Loop



David Mumford. Neuronal architectures for pattern-theoretic problems. In *Large Scale Neuronal Theories of the Brain*, 125–152. MIT Press, 1994.

Reconstruction (Residual) Error



$$A = \{a_{ij}\}$$

$$B = \{b_{ij}\}$$

Analysis-Synthesis Iteration

Analysis step: solve for B holding A constant

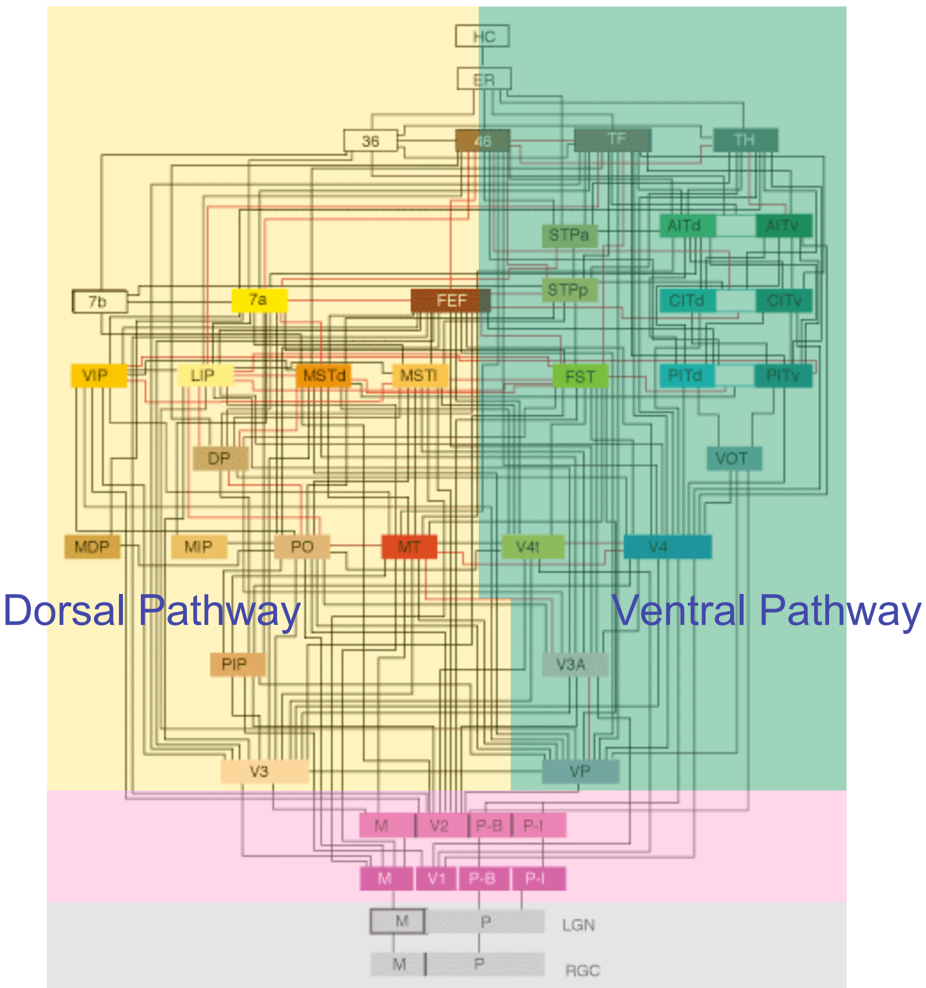
$$\text{minimize}_B\ J(B|A) = \|X - AB\|_2^2$$

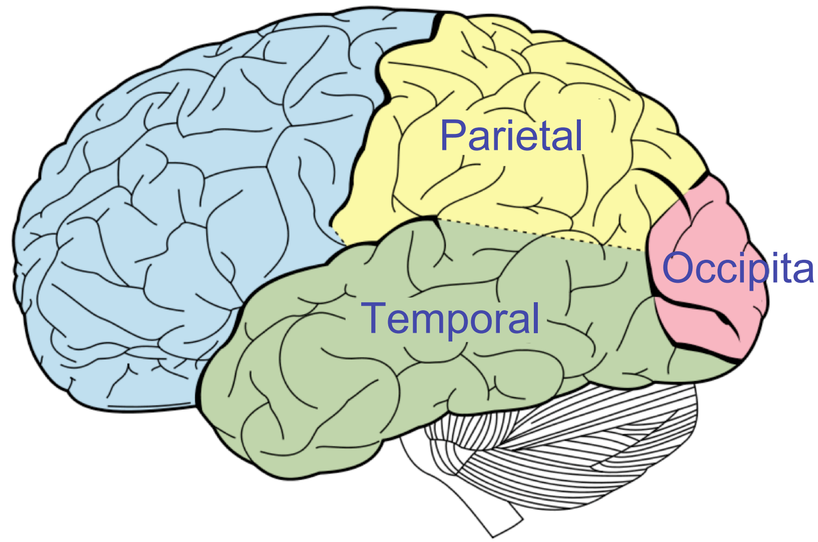
subject to $\sum_{i=1}^{L} B_{i,j}^2 < c$

Synthesis step: solve for A holding B constant

$$\text{minimize}_A\ J(A|B) = \|X - AB\|_2^2 + \lambda\|A\|_1$$

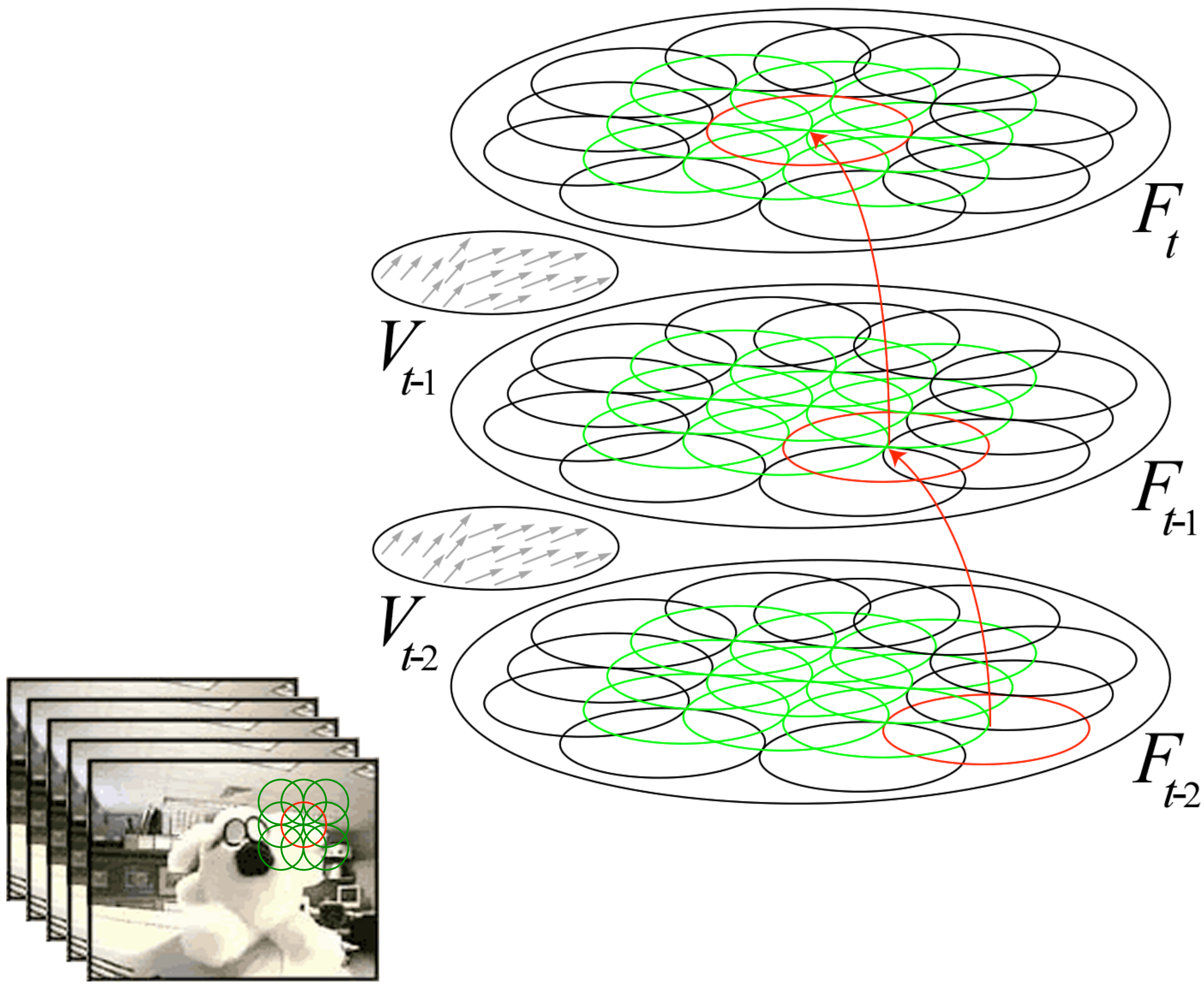
Hierarchical Graphical Models



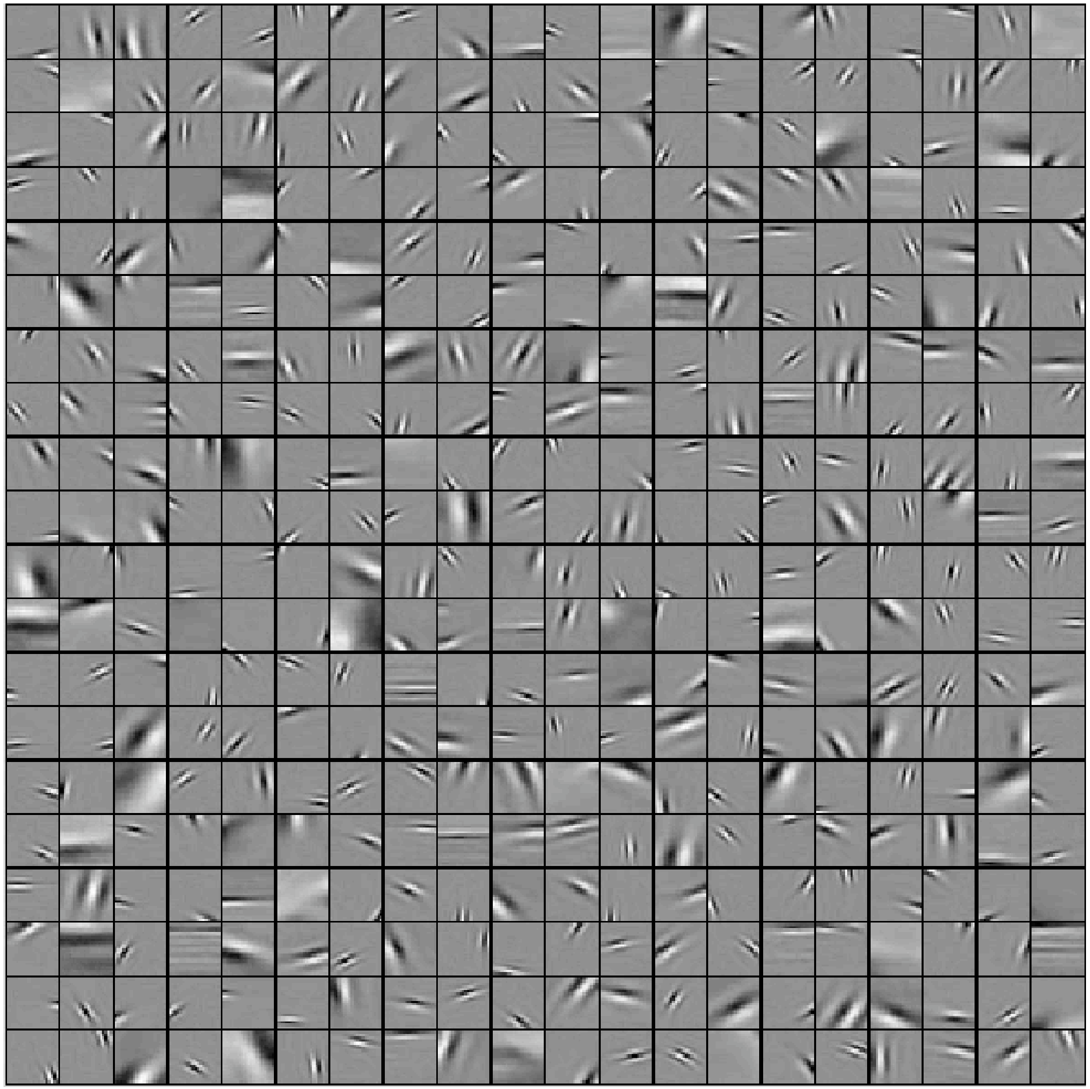


Felleman, D. J. and Van Essen, D. C. *Distributed hierarchical processing in primate cerebral cortex*. Cerebral Cortex, 1:1-47, 1991.

Space-Time Features



Spatiotemporal Features



Bruno Olshausen and Charles Cadieu. Learning invariant and variant components of time-varying natural images. *Journal of Vision.* 7(9):964–964, 2007.

Tasks of Biological Vision

- Recognition
  – Faces, predators, landmarks, *etc*.
- Navigation
  – Localizing, orienting, pursuing, *etc*.
- Manipulation
  – Grasping, throwing, assembling, *etc*.

J. Aloimonos, A. Bandyopadhyay and I. Weiss. *Active vision*. In the Proceedings of the First International Conference on Computer Vision, pages 35–55, 1987.

Ruzena Bajcsy. *Active perception*. Proceedings of the IEEE, 76(8):966–1005, 1988.

James J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, 1979.

Caltech 101 & 256 *Too Easy?*

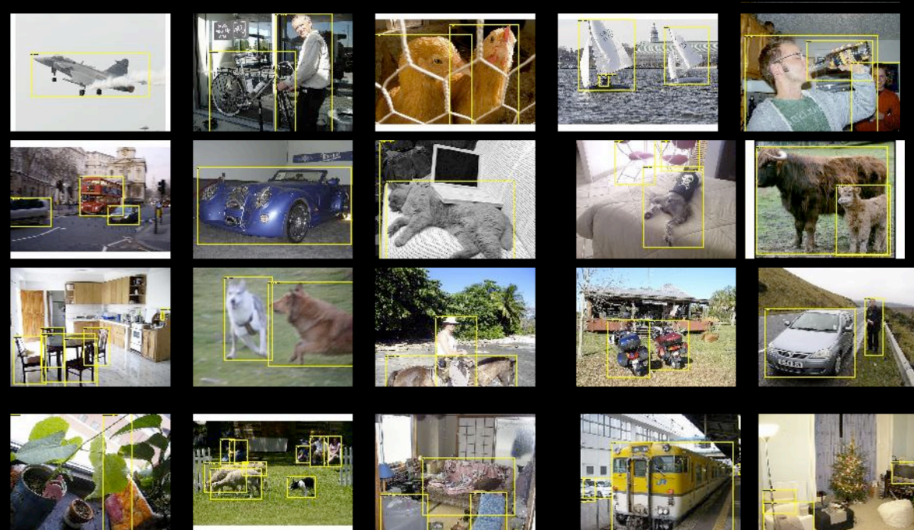
Fei-Fei, Fergus, Perona, 2004

Griffin, Holub, Perona, 2007

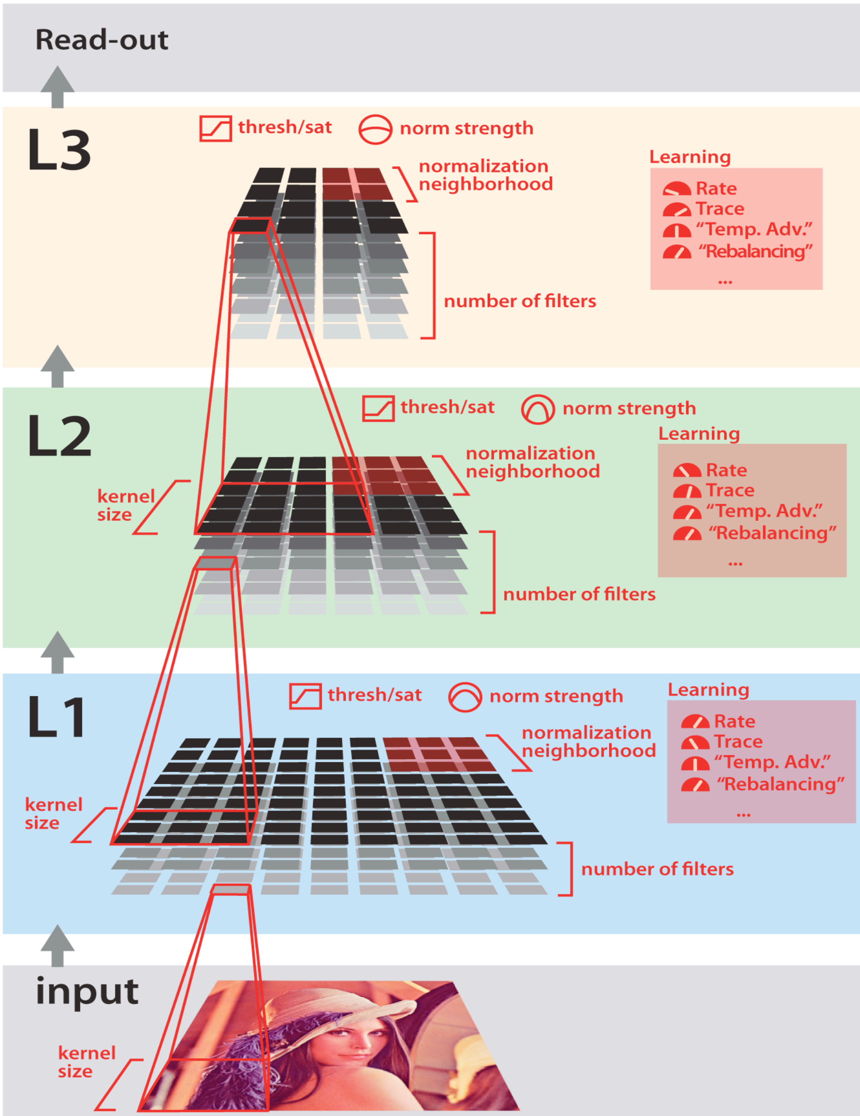
The PASCAL Visual Object Classes Challenge 2007 *Too Hard?*

The twenty object classes that have been selected are:

Person: person
Animal: bird, cat, cow, dog, horse, sheep
Vehicle: aeroplane, bicycle, boat, bus, car, motorbike, train
Indoor: bottle, chair, dining table, potted plant, sofa, tv/monitor



M. Everingham, Luc van Gool , C. Williams, J. Winn, A. Zisserman 2007



Generate Thousands of Random Models

↓

Unsupervised Learning of all models *(within a given "petri dish")*

↓

Test all models with a "screening" object recognition task

↓

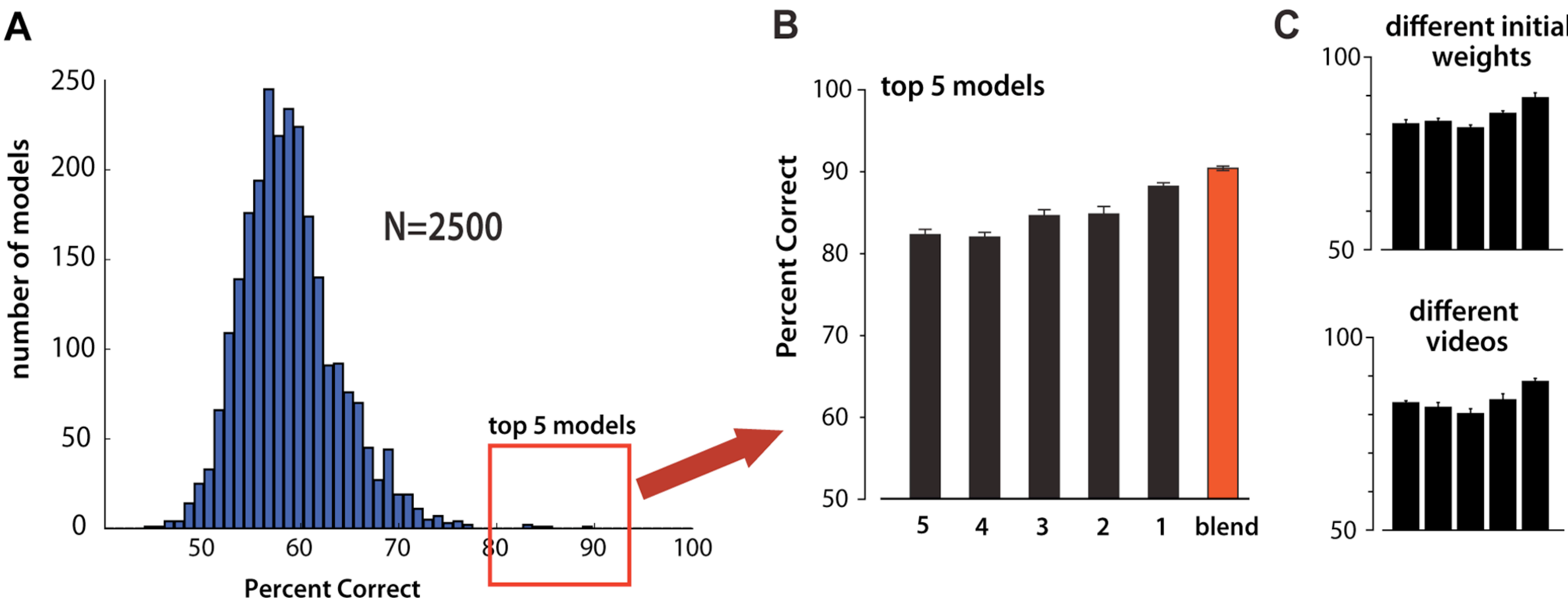
Choose the best models

↓

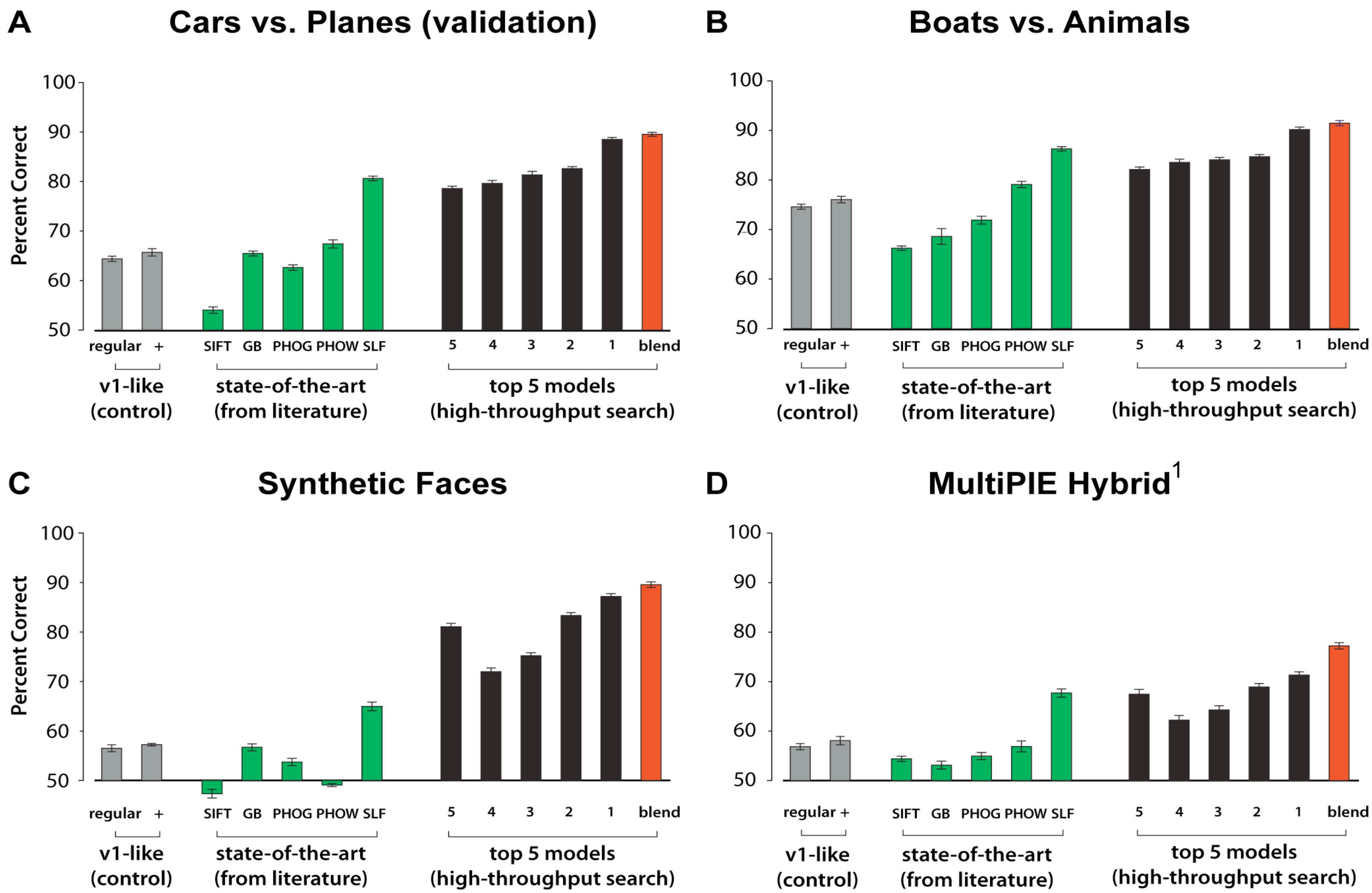
Validate on other tasks

Nicolas Pinto, David Doukhan, James DiCarlo, and David Cox. A high-throughput screening approach to discovering good forms of biologically inspired visual representation. *PLoS Computational Biology*, 5(11):e1000579, November 2009.

Searching for Top Performing Models in the Long Tail



Nicolas Pinto, David Doukhan, James DiCarlo, and David Cox. A high-throughput screening approach to discovering good forms of biologically inspired visual representation. *PLoS Computational Biology*, 5(11):e1000579, November 2009.

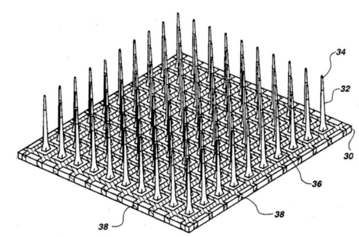


SIFT: Scale Invariant Feature Transform [Lowe, 2004]
GB: Geometric Blur [Berg & Malik, 2001]
PHOG: Pyramidal Histogram of Gradients [Lazebnik *et al*, 2006]
PHOW: Pyramidal Histogram of Words [Bosch *et al*, 2007]
SLF: Sparse Localized Features [Mutch & Lowe, 2008]

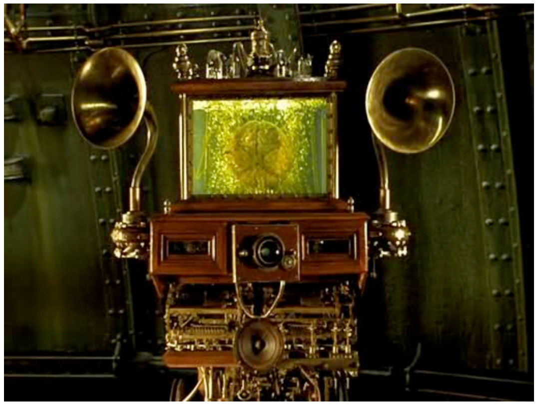
1. Face Multiple Pose Illumination Expression Dataset

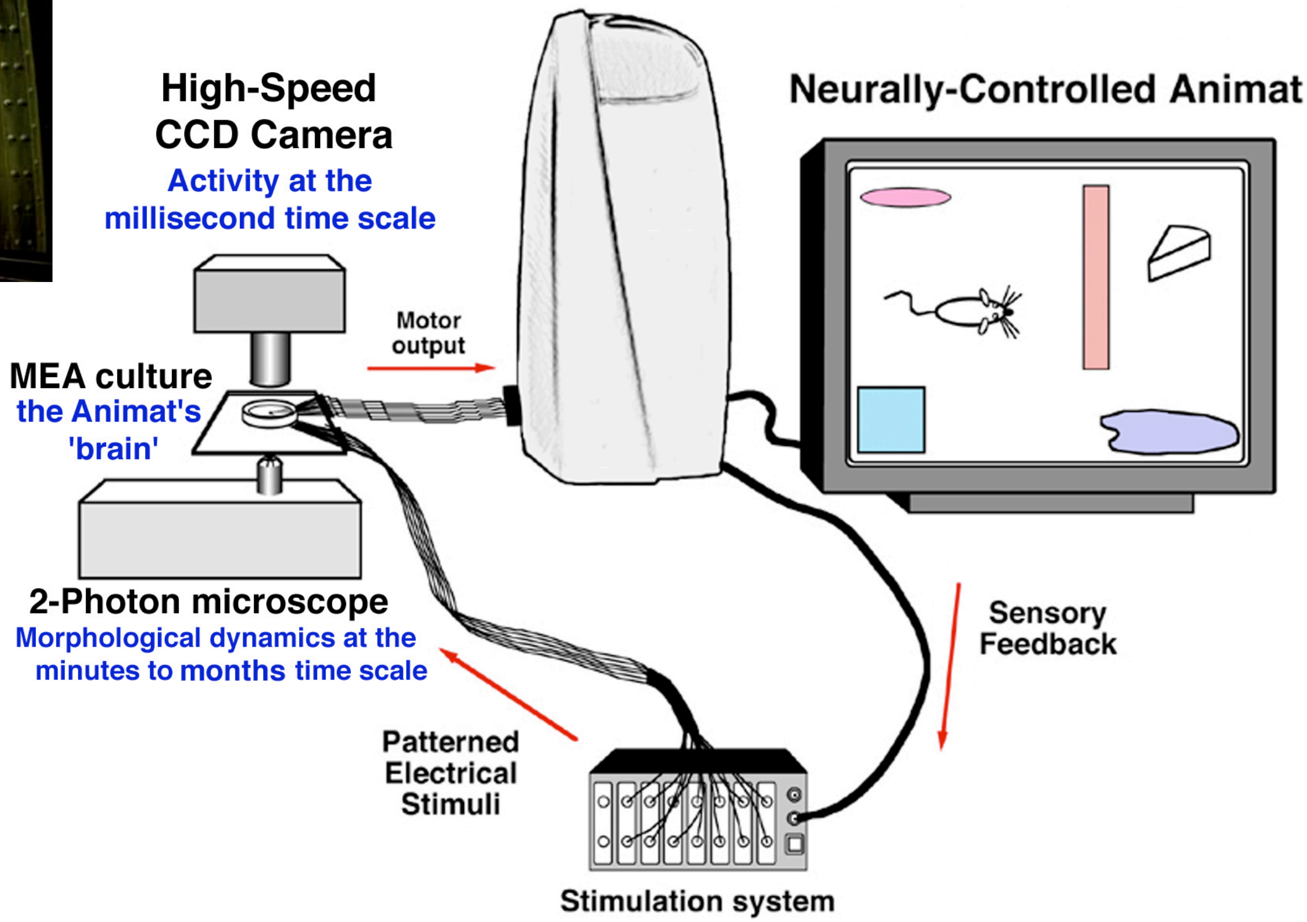
Imaging, Recording Activity and Stimulating Cortical Tissue

- Positron Emission Tomography (PET)
- Functional Magnetic Resonance Imagery
- Single Neuron Spike Recordings
- Multielectrode Implantable Arrays
- Two-Photon Microscope Imaging
- Magnetoelectroencephalography (MEG)
- Transcranial magnetic stimulation (TMS)



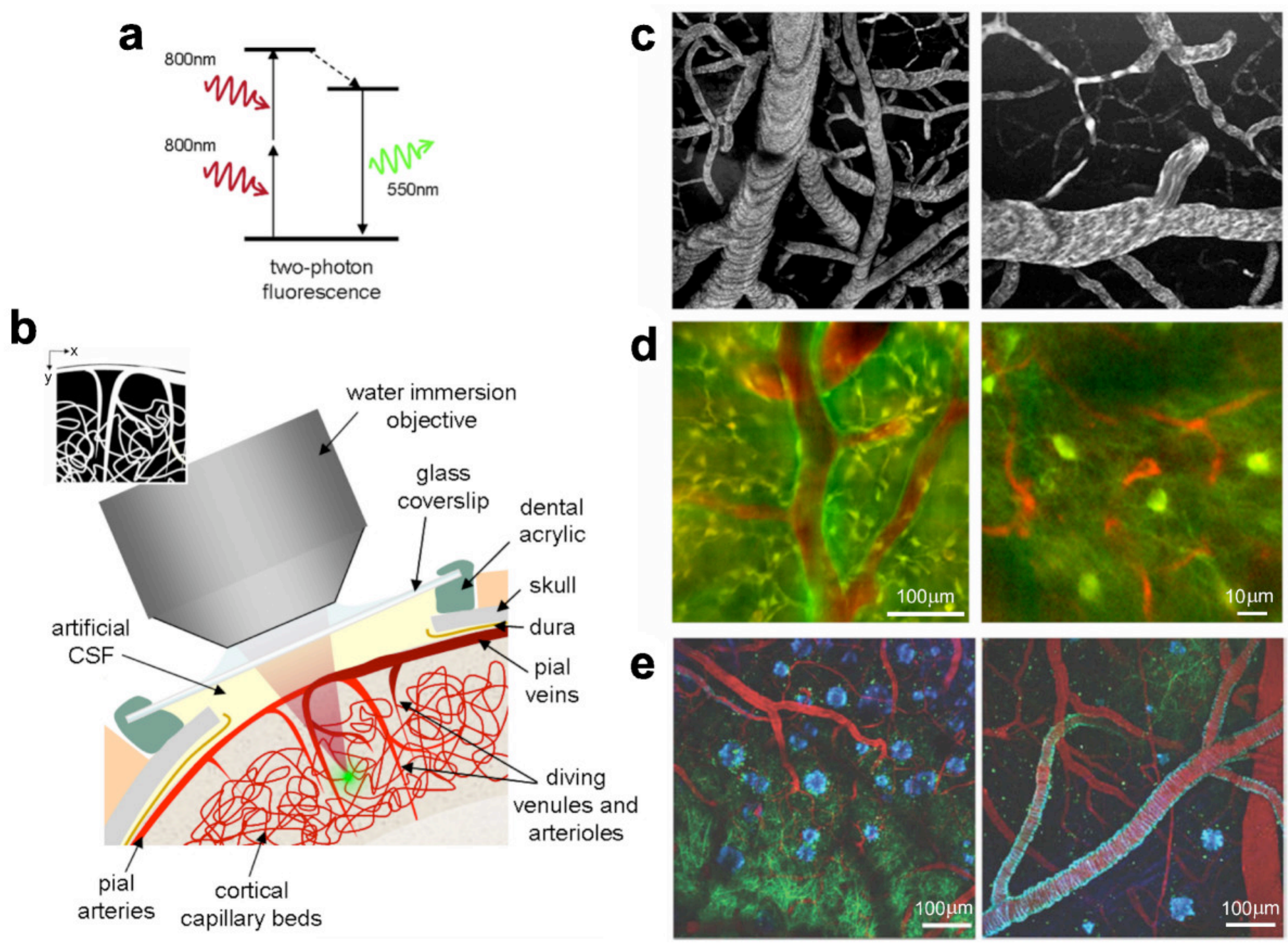
Brain in a Vat *in vitro* Experiment

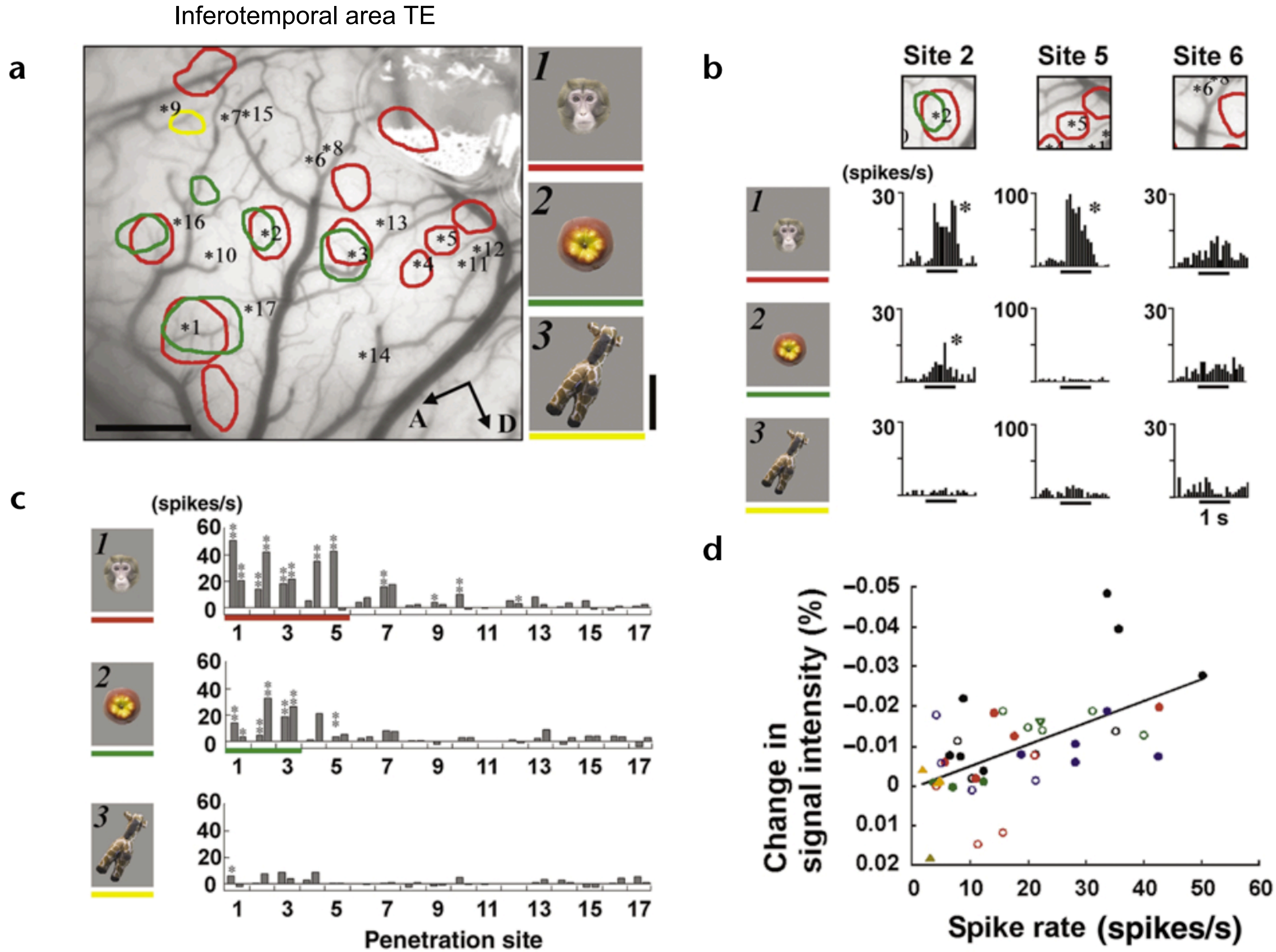




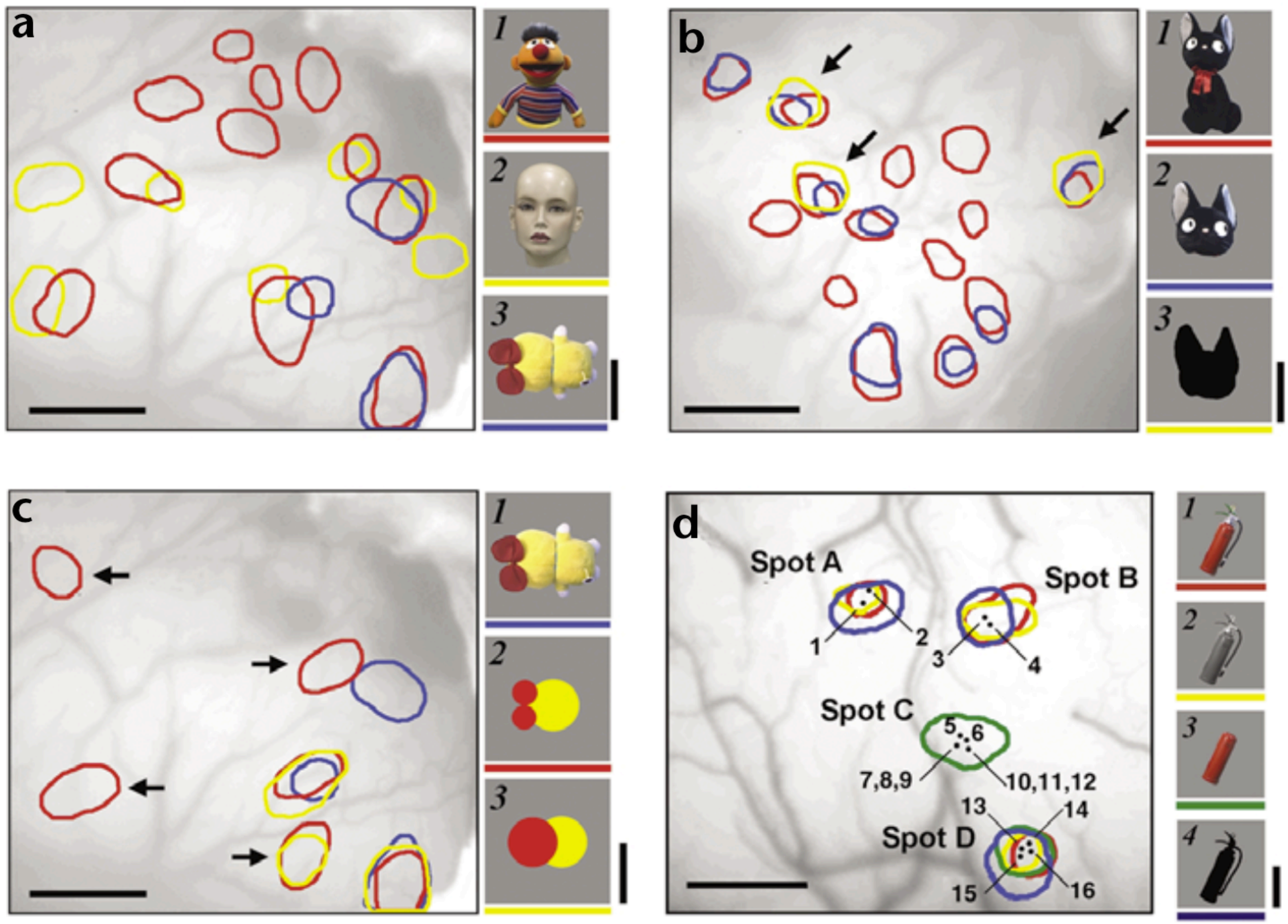
Thomas B. DeMarse, Daniel A. Wagenaar, Axel W. Blau, and Steve M. Potter. The neurally controlled animat: Biological brains acting with simulated bodies. *Autonomous Robots*, 11(3):1573–7527, 2001.

In Vivo Two-Photon Microscopy





Y. Yamane, K. Tsunoda, M. Matsumoto, N. A. Phillips, and M. Tanifuji. Representation of the spatial relationship among object parts by neurons in macaque inferotemporal cortex. *Journal Neurophysiology*, 96:3147–3156, 2006.
K. Tsunoda, Y. Yamane, M. Nishizaki, and M. Tanifuji. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience*, 4:832–838, 2001.



Y. Yamane, K. Tsunoda, M. Matsumoto, N. A. Phillips, and M. Tanifuji. Representation of the spatial relationship among object parts by neurons in macaque inferotemporal cortex. *Journal Neurophysiology*, 96:3147–3156, 2006.
K. Tsunoda, Y. Yamane, M. Nishizaki, and M. Tanifuji. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience*, 4:832–838, 2001.

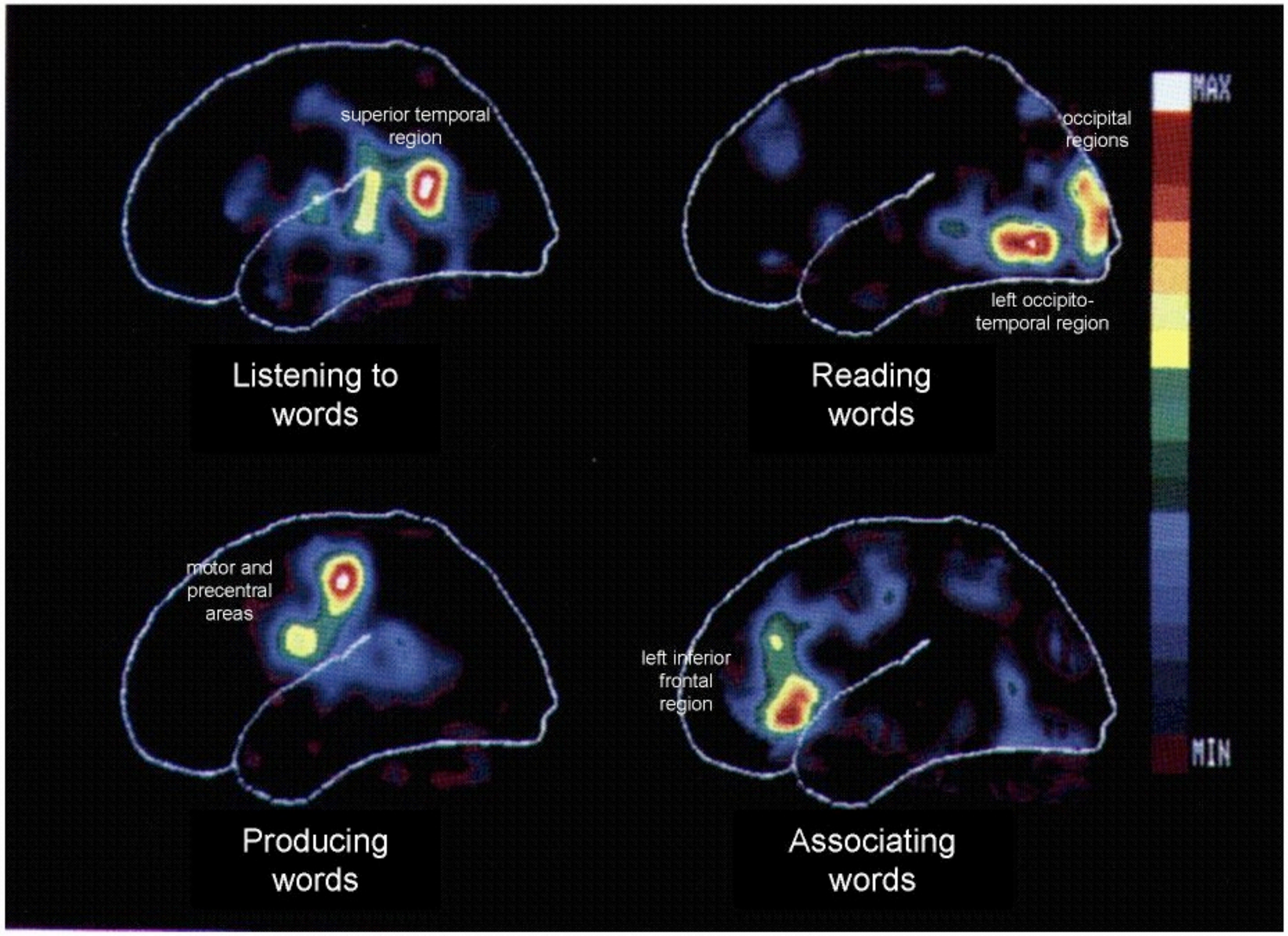


Figure 2.3. A historical image: the brain areas for language, as first revealed by PET scanning (data from Petersen et al., 1989; image courtesy of Marcus Raichle). Relative to the fixation of a small dot, silent reading (top right) activates processes of visual word recognition located in the rear part of the left hemisphere. Depending on the task, information is then transmitted to regions coding for speech sounds (top left), speech production (bottom left), or the manipulation of word meanings (bottom right).

Stanislas Dehaene. *Reading in the Brain: The Science and Evolution of a Human Invention*. Viking Press, 2009.

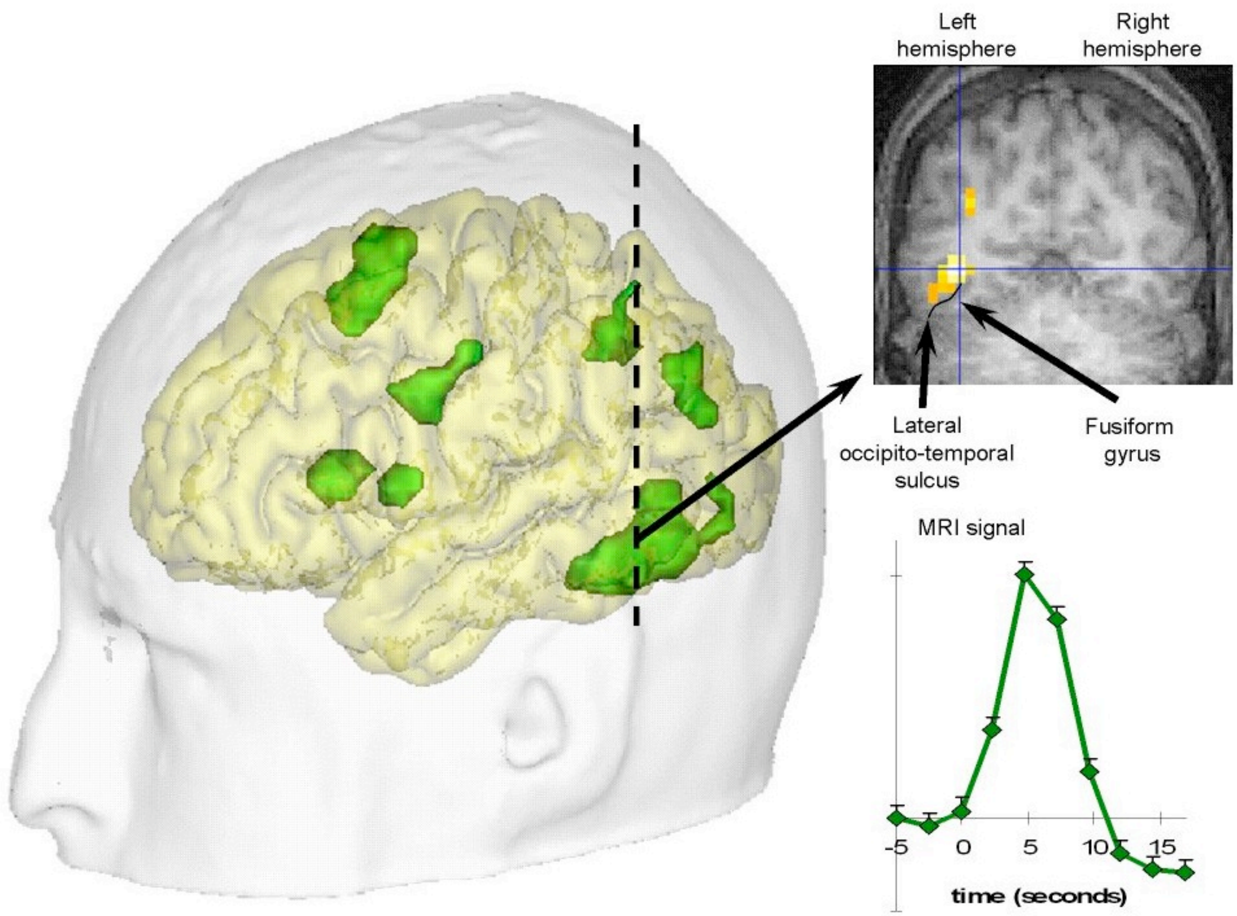
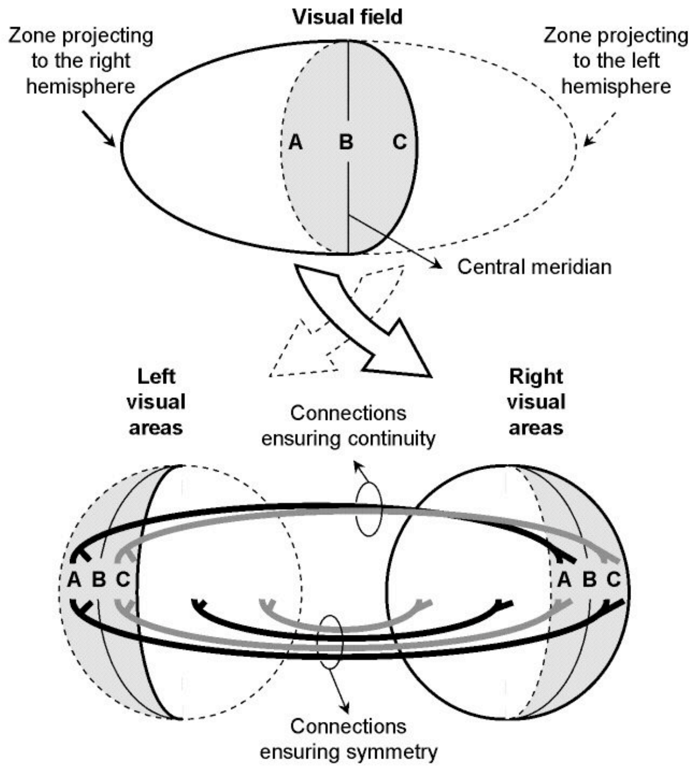
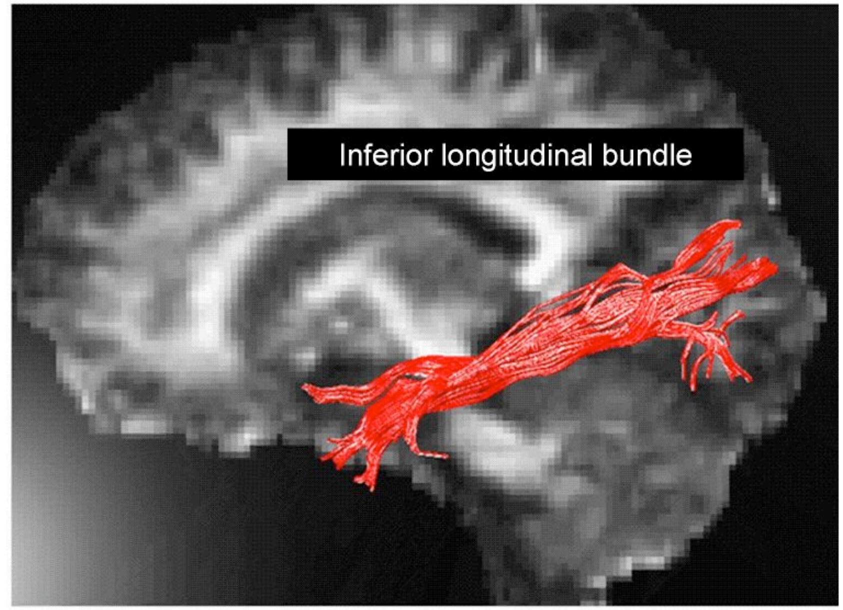
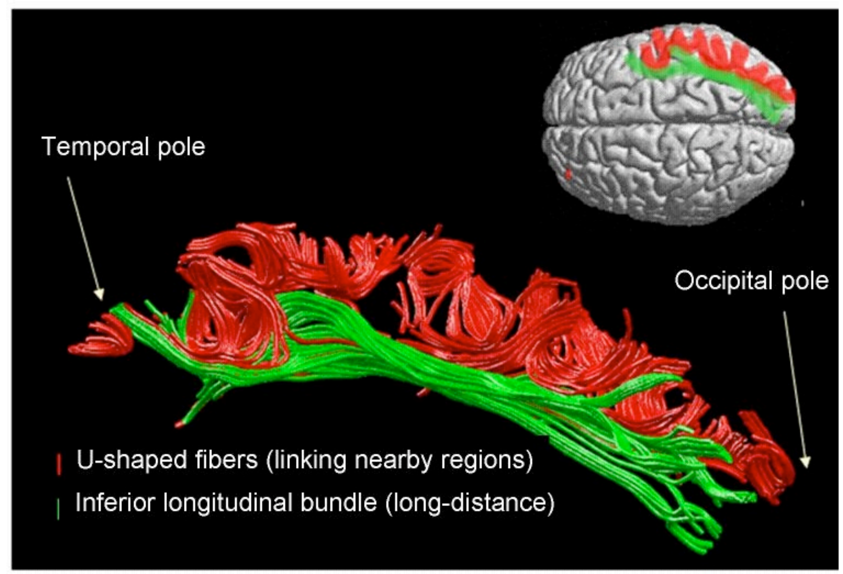


Figure 2.4. Functional magnetic resonance imaging (fMRI) can locate the brain areas involved in reading in just a few minutes. Participants read words presented at random intervals. After each word, reading areas show a characteristic increase in MRI signal which reaches a peak about five seconds later. The active network varies depending on the exact task and the nature of the control state. However, it always includes the visual word form area, the "brain's letterbox". This region is systematically located deep in the left lateral occipito-temporal sulcus, next to the fusiform gyrus.

Stanislas Dehaene. *Reading in the Brain: The Science and Evolution of a Human Invention*. Viking Press, 2009.





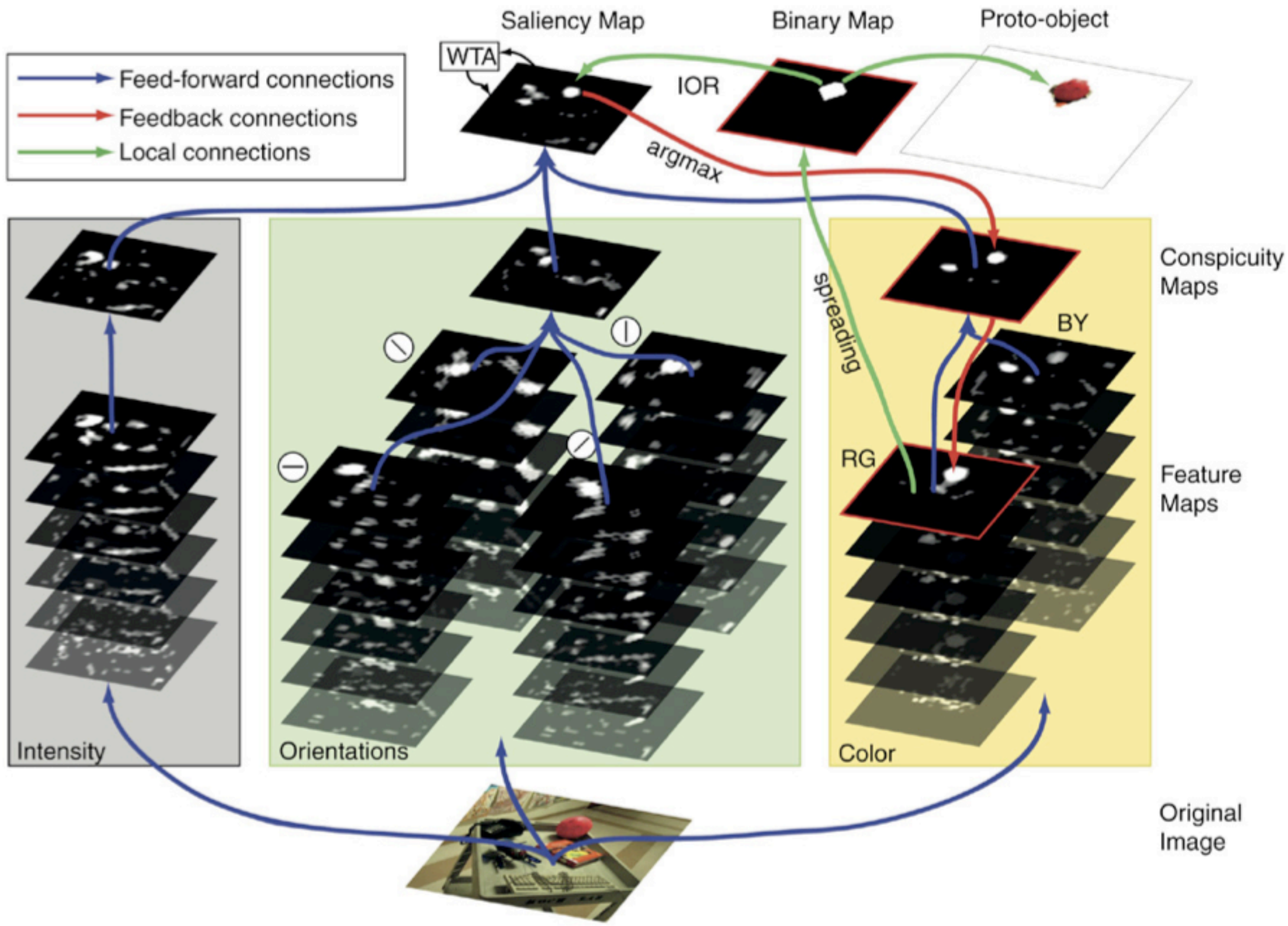


Stanislas Dehaene. *Reading in the Brain: The Science and Evolution of a Human Invention*. Viking Press, 2009.

Tyranny of Retinotopic Maps

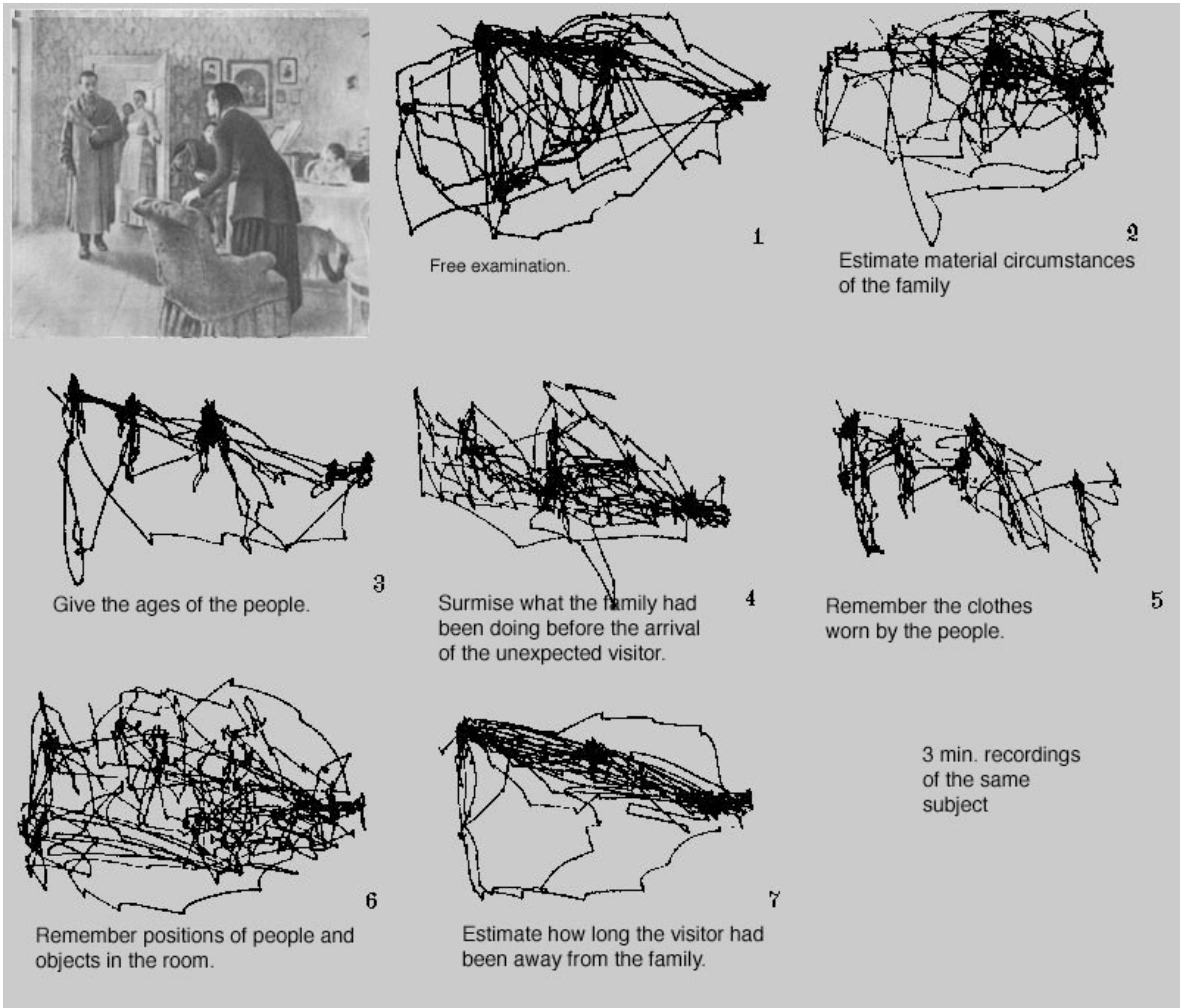
- Images on the retina are ephemeral
  - Saccades, ego and other motion, *etc*.
- Integrating these fleeting glimpses
  - Stitching, remembering, inventing, *etc*.
- Moving between coordinate frames
  - Our head, body, car, bicycle, *etc*.
  - Other people, objects, locations, *etc*.

Attention and Saliency



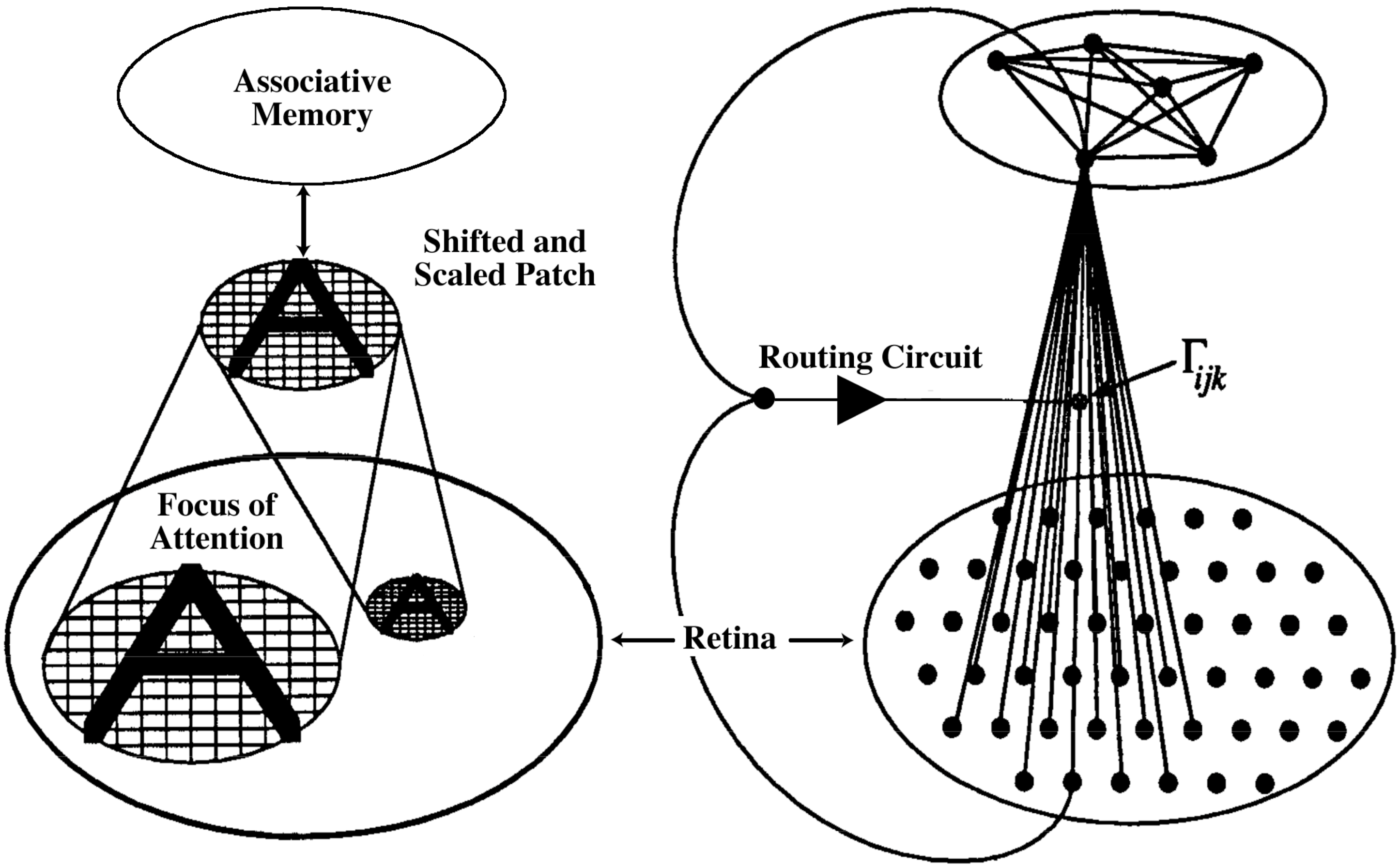
Laurent Itti and Christof Koch. *A saliency-based search mechanism for overt and covert shifts of visual attention*. Vision Research 40(10–12):1489–1506, 2000.

Dirk Walther and Christof Koch. *Modeling attention to salient proto-objects*. Neural Networks 19:1395-1407, 2006.



Yarbus, A. L. *Eye Movements and Vision*. Plenum. New York. 1967 (Originally published in Russian 1962)

Routing Circuit Model of Invariant Pattern Recognition

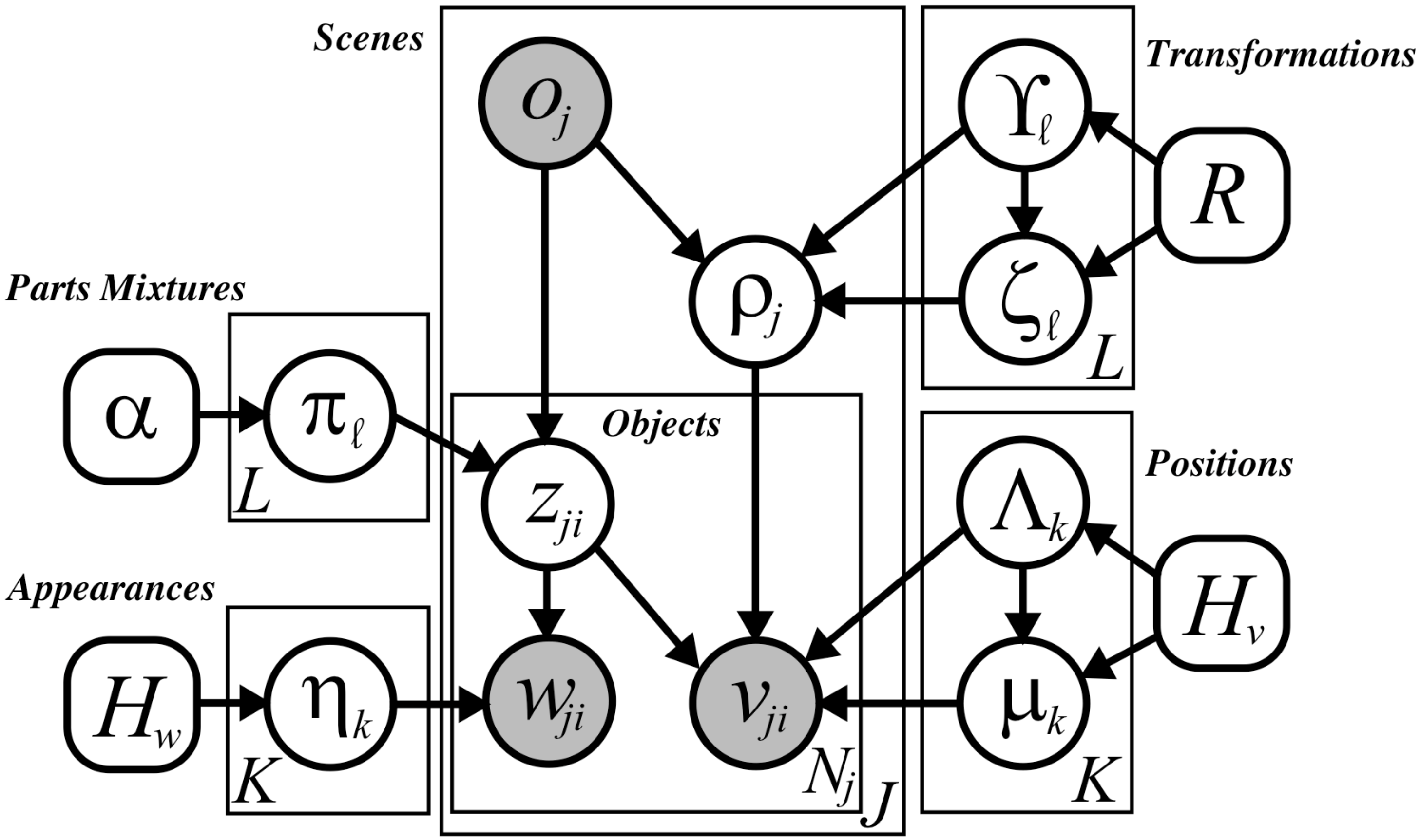


Bruno Olshausen, Charles Anderson and David Van Essen. A Neurobiological Model of Visual Attention and Invariant Pattern Recognition Based on Dynamic Routing of Information. The Journal of Neuroscience, November 1993, 13(11): 4700-4719.
Geoffrey F. Hinton. A parallel computation that assigns canonical object-based frames of reference. In *Proceedings of the 7th International Joint Conference on Artificial intelligence*, pages 683−685, 1981. Morgan Kaufmann.

Models and Metaphors

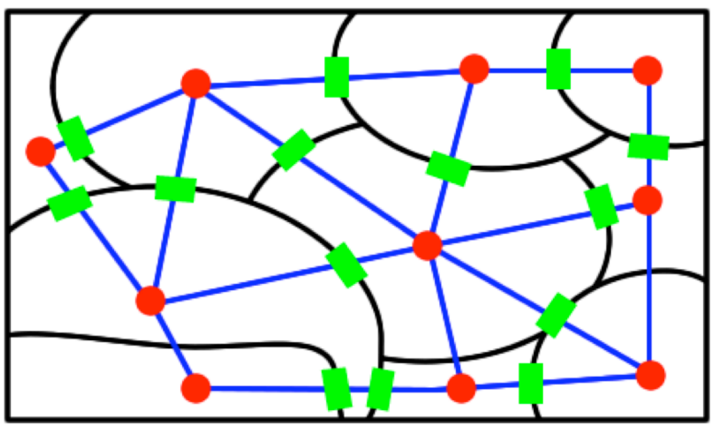
- Hydraulics — pipes, pumps, valves
- Clocks — gears, ratchets, springs
- Automata — logic circuits, computers
- Linear systems — filters, vector spaces
- Probabilistic graphical models —
  - Graphs mirror biological networks
  - Compact distributed representations
  - Computation using message passing

Modeling Shared Features with Transformed Dirichlet Processes

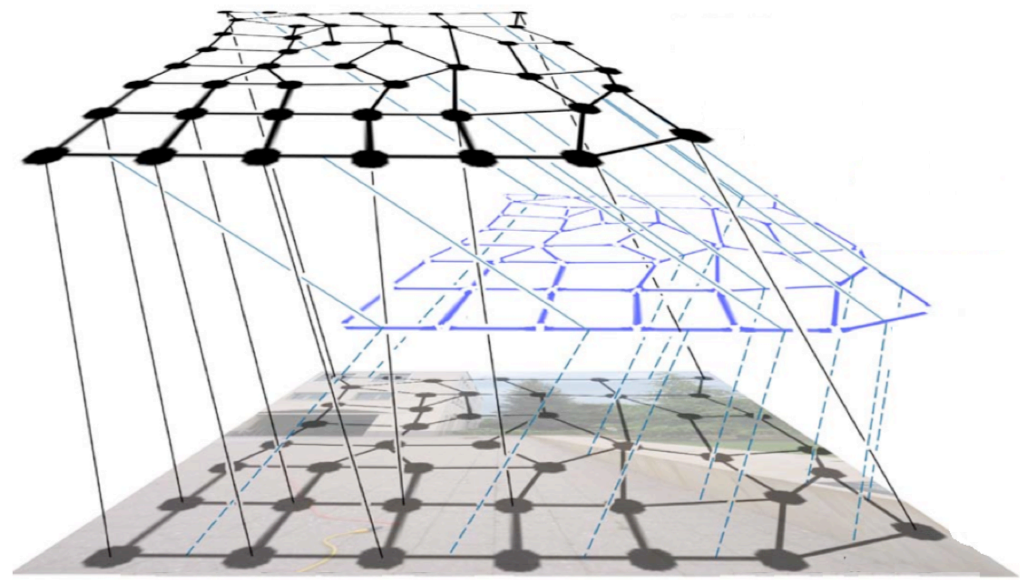


Erik B. Sudderth, Antonio Torralba, William T. Freeman, and Alan S. Willsky. Describing visual scenes using transformed objects and parts. *International Journal of Computer Vision*, 77(1-3):291–330, 2008.

Modeling Spatial Layout with Markov Random Fields



$$P(\mathbf{e} \mid \mathbf{x}_t) = \frac{1}{Z} \prod_i \varphi(e_i) \prod_j \phi(\mathbf{c}_j)$$

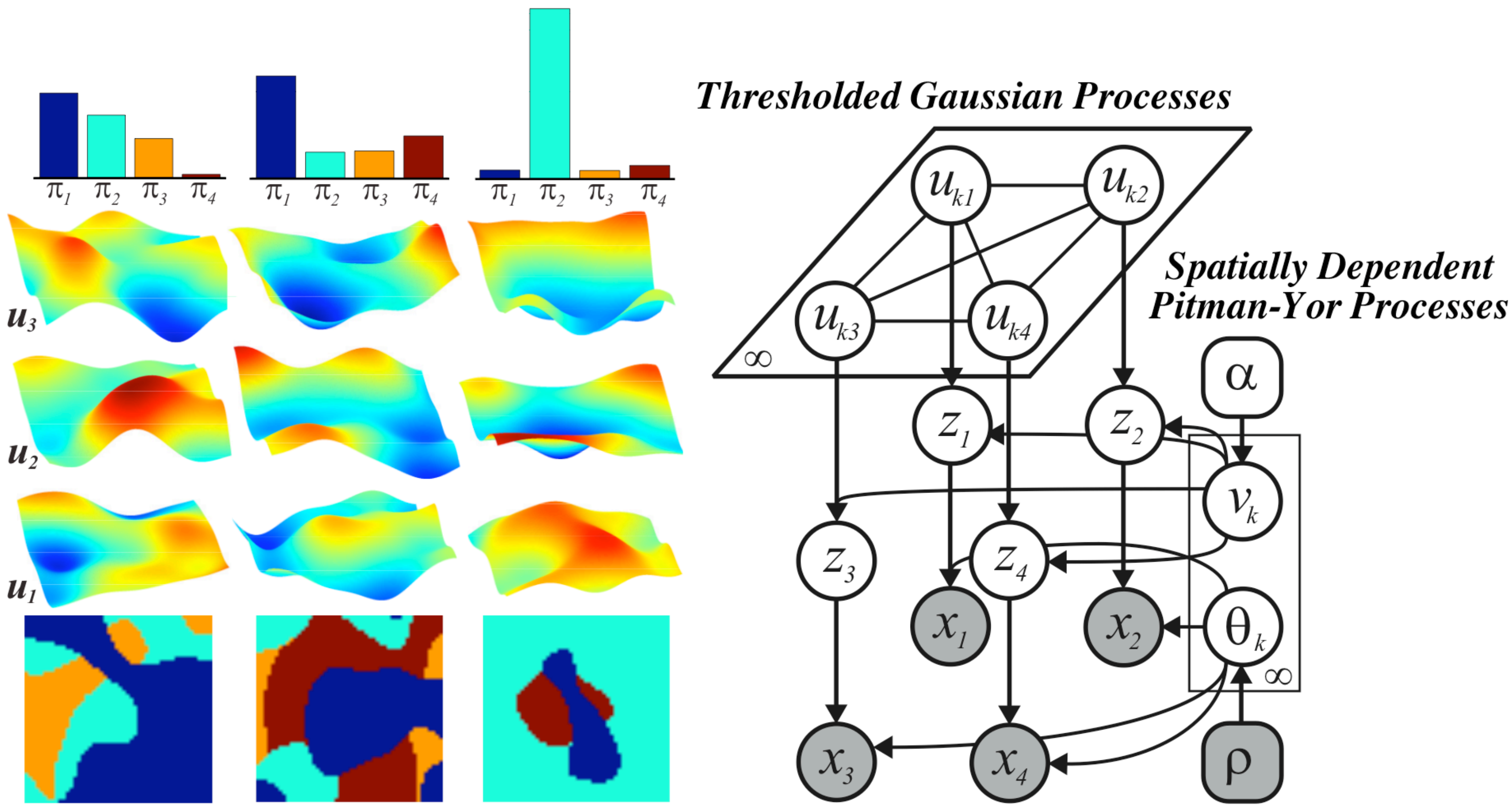


Derek Hoiem, Alexei Efros, and Martial Hebert. Putting objects in perspective. *Conference on Computer Vision and Pattern Recognition*, IEEE, 2137−2144, 2006.

Ashutosh Saxena, Sung Chung, and Andrew Ng. 3-D depth reconstruction from a single still image. *International Journal of Computer Vision*, 76(1):53−69, 2007.

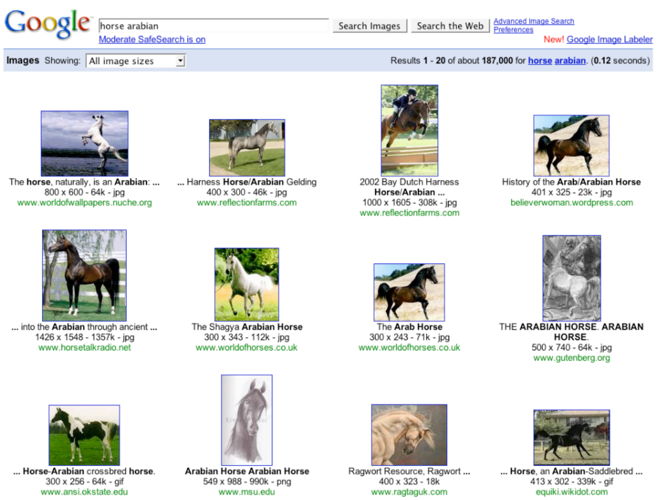
Stephen. Gould, Richard Fulton, and Daphne Koller. Decomposing a scene into geometric and semantically consistent regions. *International Conference on Computer Vision,* 2009.

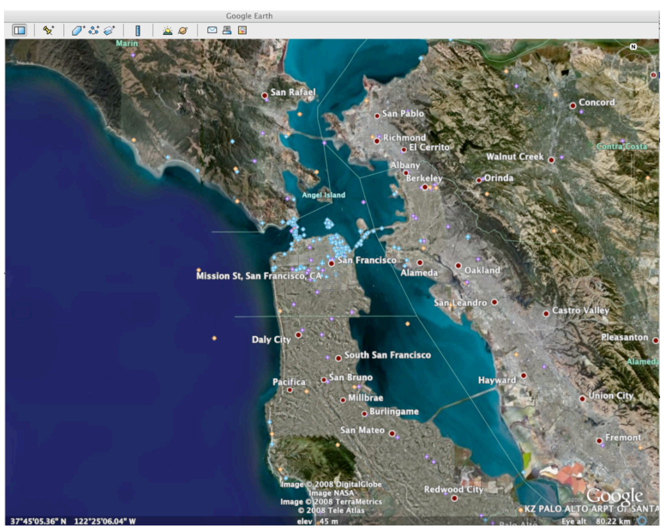
Modeling Spatial Layout with Dependent Pitman-Yor Processes

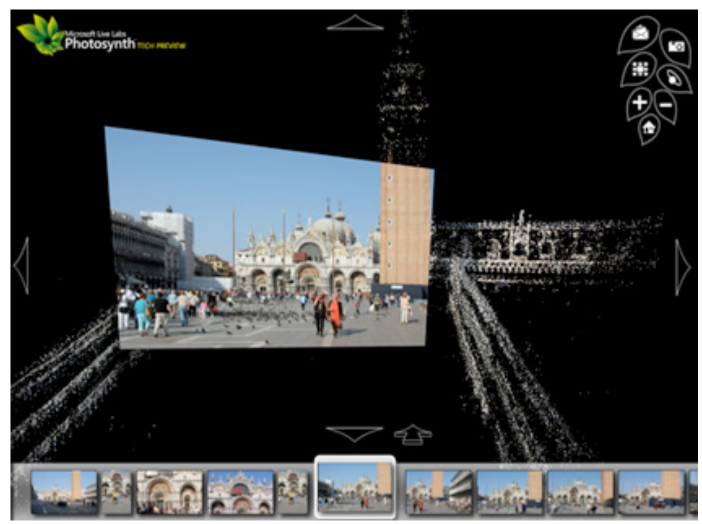


Erik Sudderth and Michael Jordan. Shared segmentation of natural scenes using dependent Pitman-Yor processes. In *Advances in Neural Information Processing Systems 21*, 1585–1592. 2009.

World Wide Visual Memory







World Wide Visual Memory

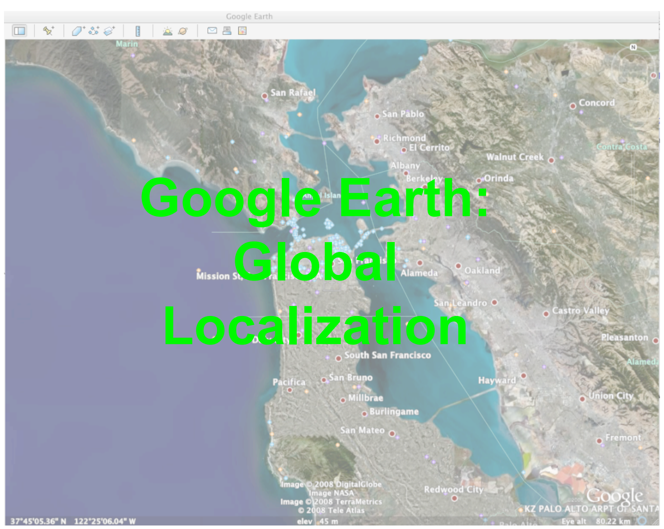
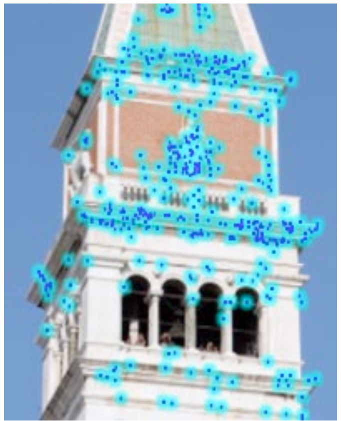


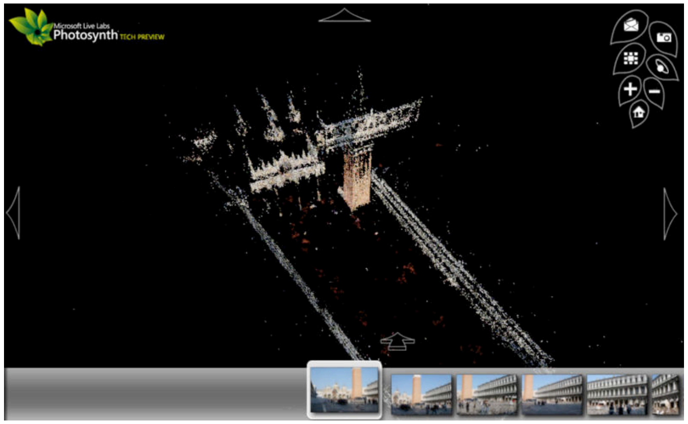
Image Stitching Writ Large

Interest Points

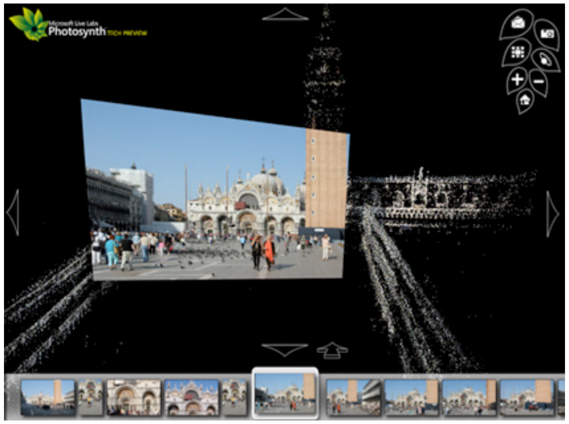


Rectification

Point Clouds

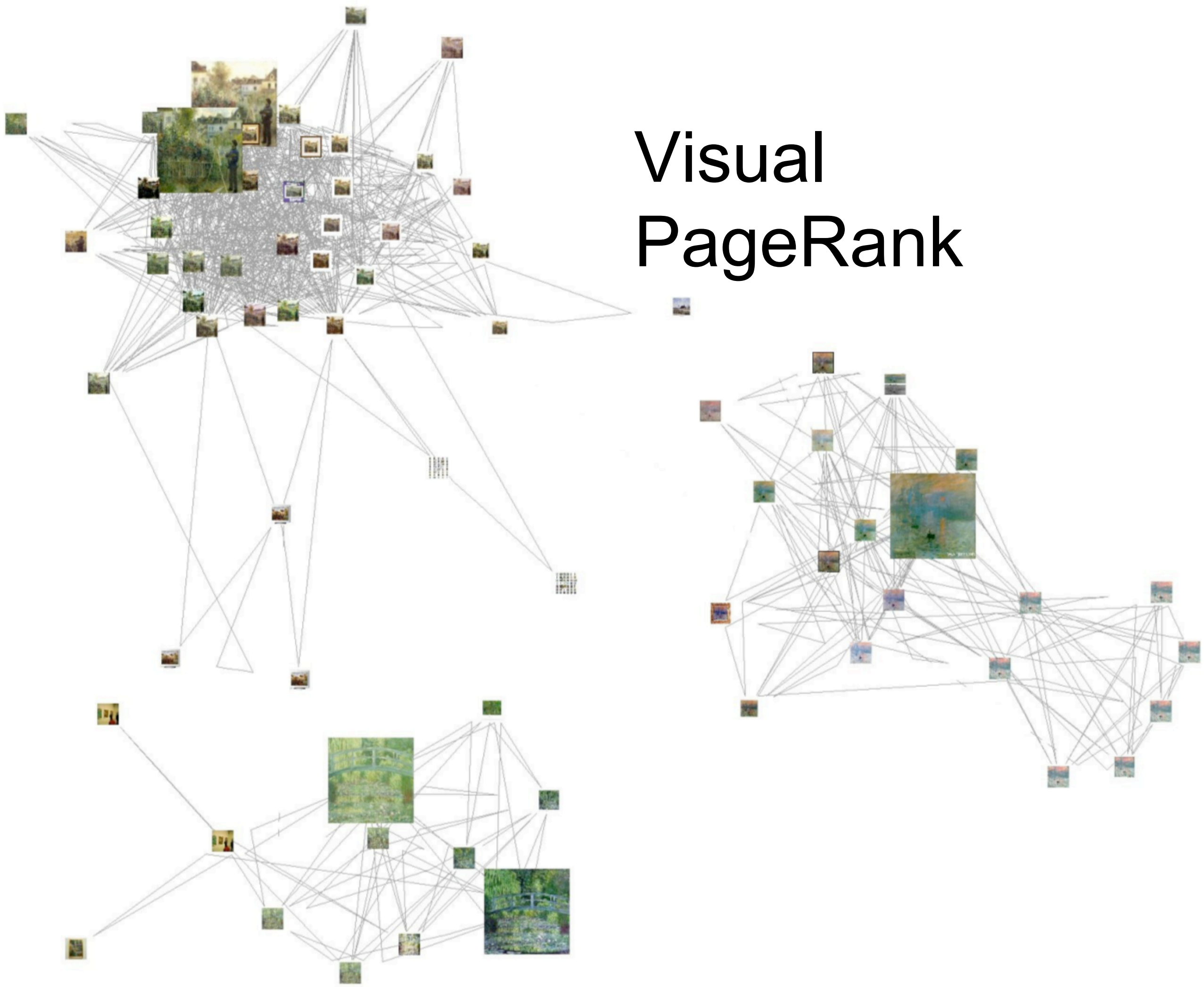


Interpolation



Parallel Graph Algorithms

Visual
PageRank



Jing, Yushi and Baluja, Shumeet. *PageRank for Product Image Search*. In Proceedings of the 17th World Wide Web Conference. Beijing, China, 2008.

CS379C Administrivia

- Background lectures by the instructor[1]
- Student presentations on select papers
- Project discussions led by instructor
- Class grades based on
  - Class participation including presentation (20%)
  - Project proposal due around midterm (20%)
  - Project documentation and demonstration (60%)

---

1. I've also invited several experts to come and discuss key issues; more about this as we work out the schedule of papers and coordinate with the invitees.