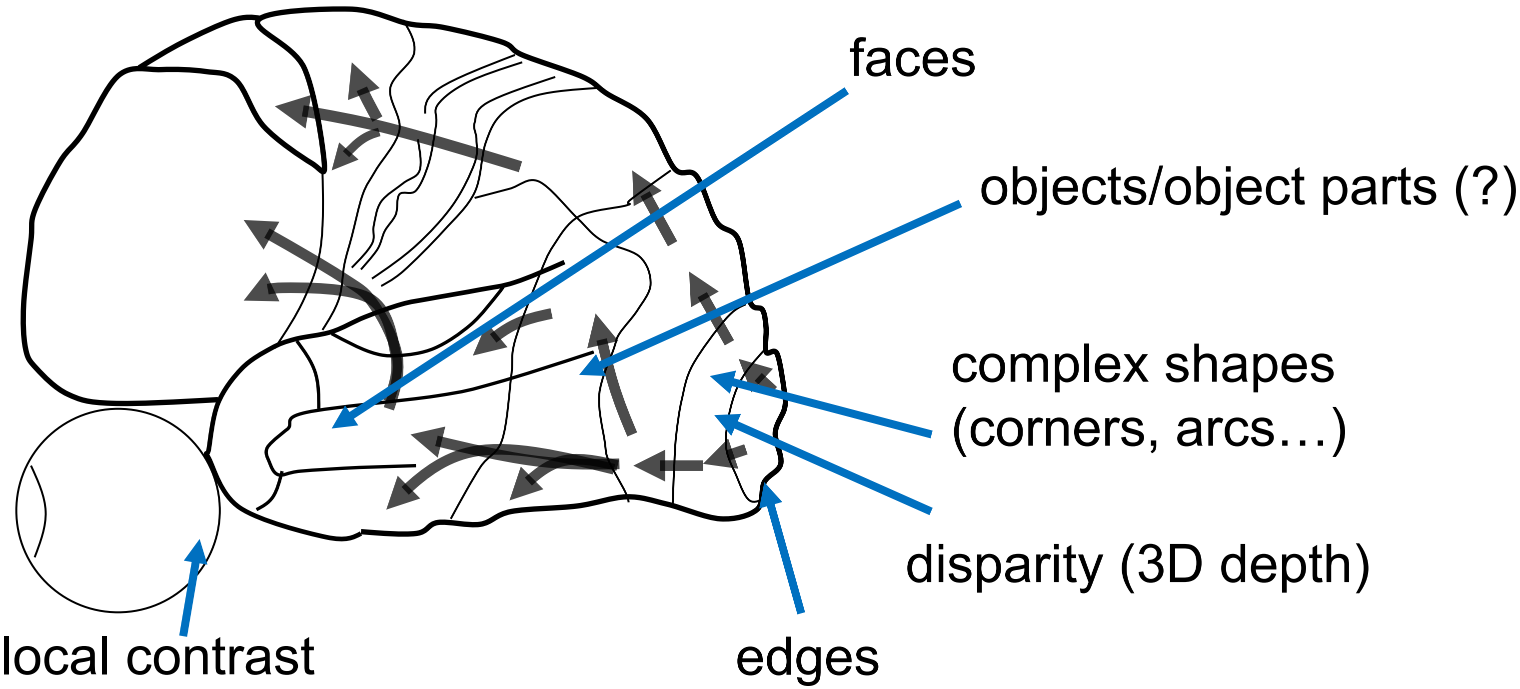
Knowledge Processing with Neural Networks

Lecture 8

Neural Representations in the Visual System

http://www.informatik.uni-hamburg.de/WTM/

Representation in the Visual System of the Brain



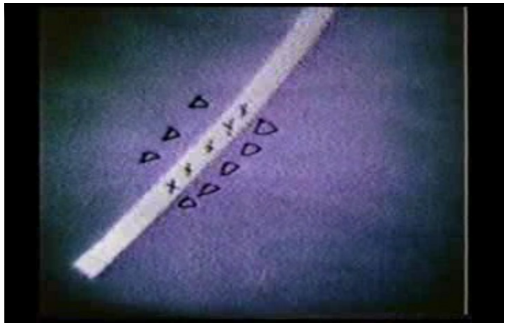
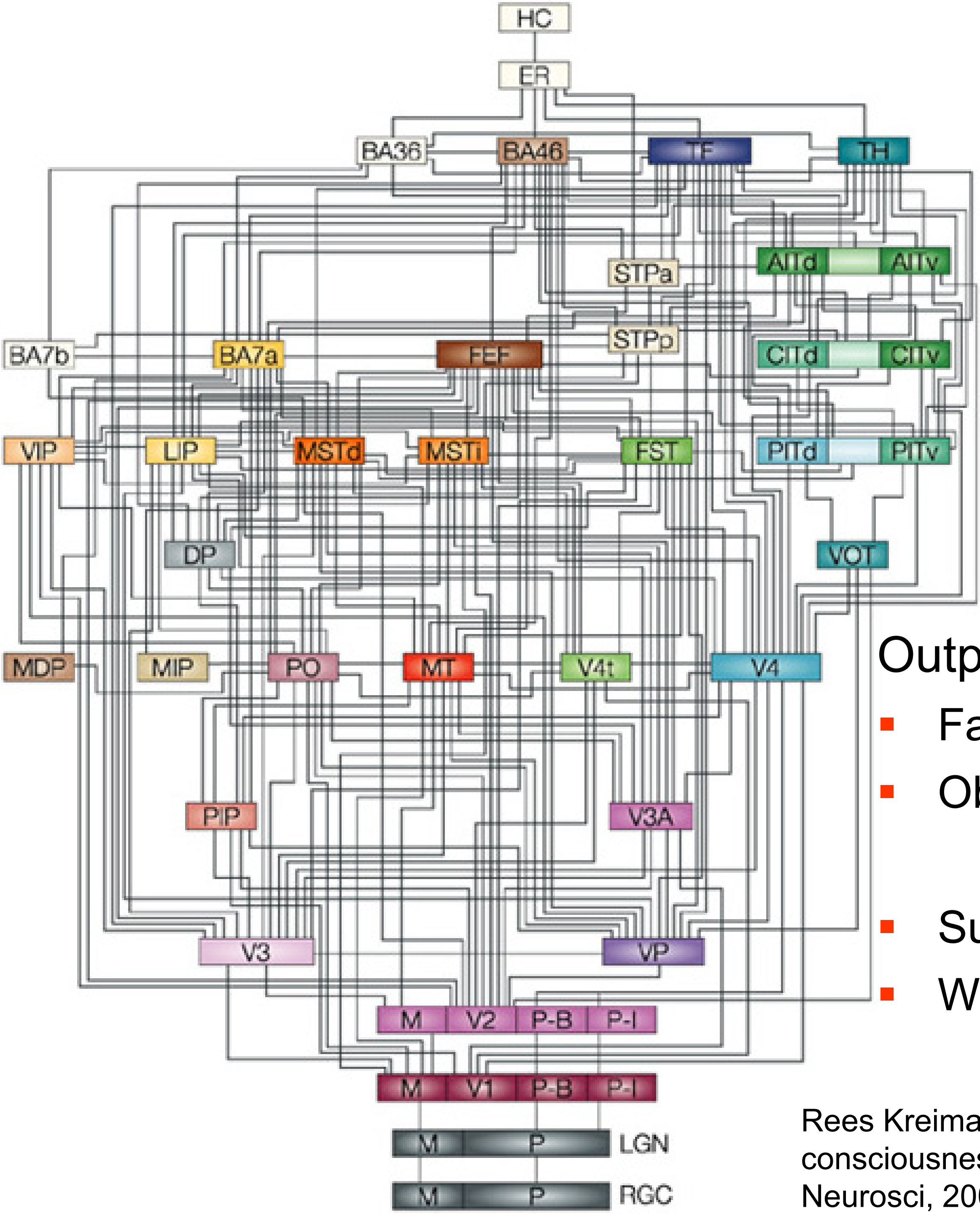


Figure from: www.benbest.com/science/anatmind/anatmd6.html
Videos by Hubel & Wiesel

Hierarchy of the Visual System



Outputs:

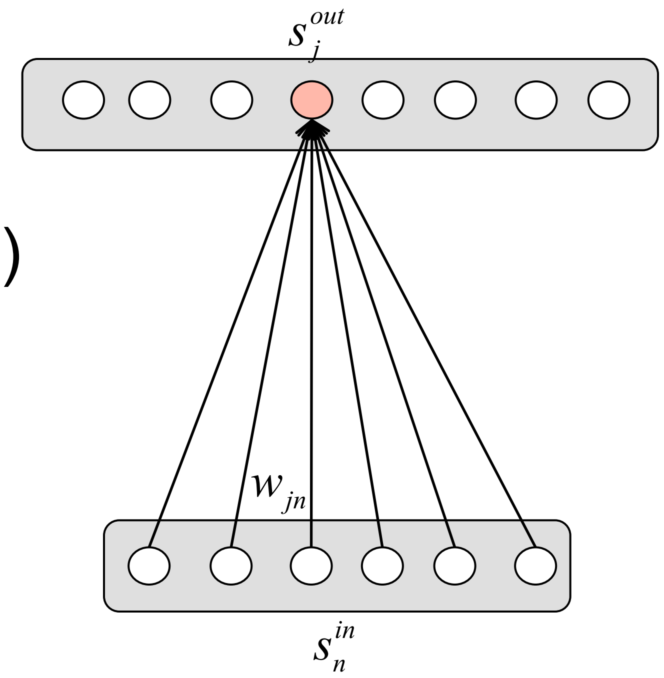
- Faces: Who, emotion, attention
- Objects: what, where,
  how to grasp
- Surround: where am I?
- Where to look next?

Rees Kreiman Koch. Neural correlates of consciousness in humans. Nature Reviews Neurosci, 2002

Recall: Perceptron/Connectionist Neurons

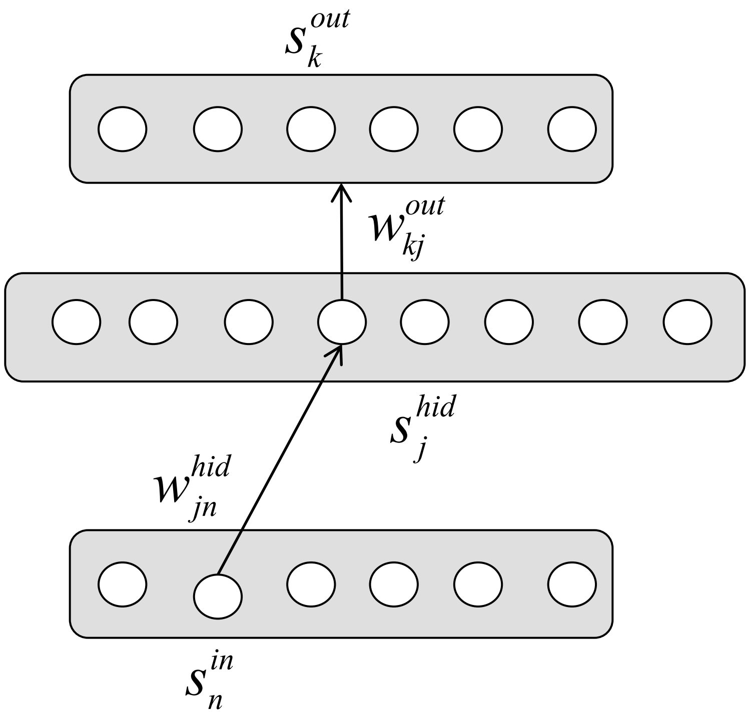
- Activate a neuron: $s_j^{out} = \sum_n w_{jn} s_n^{in} = \vec{w}_{jn} \cdot \vec{s}^{in}$
  - Dot product (scalar product) between weight vector and input vector
- Activate all neurons: $\vec{s}^{out} = W\vec{s}^{in}$
  - Matrix product with weight matrix
- In Python: s_out = numpy.dot(W,s_in)
- In C: two nested for-loops
  - Outer loop over output neurons,
    inner loop does scalar product
- Transfer function applied, e.g.:

$$y_j^{out} = \tanh(s_j^{out})$$

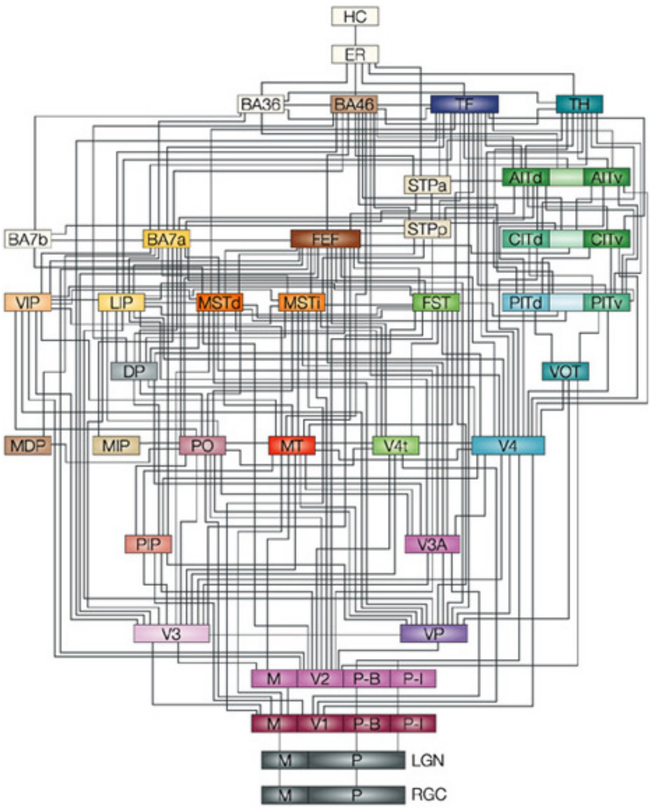


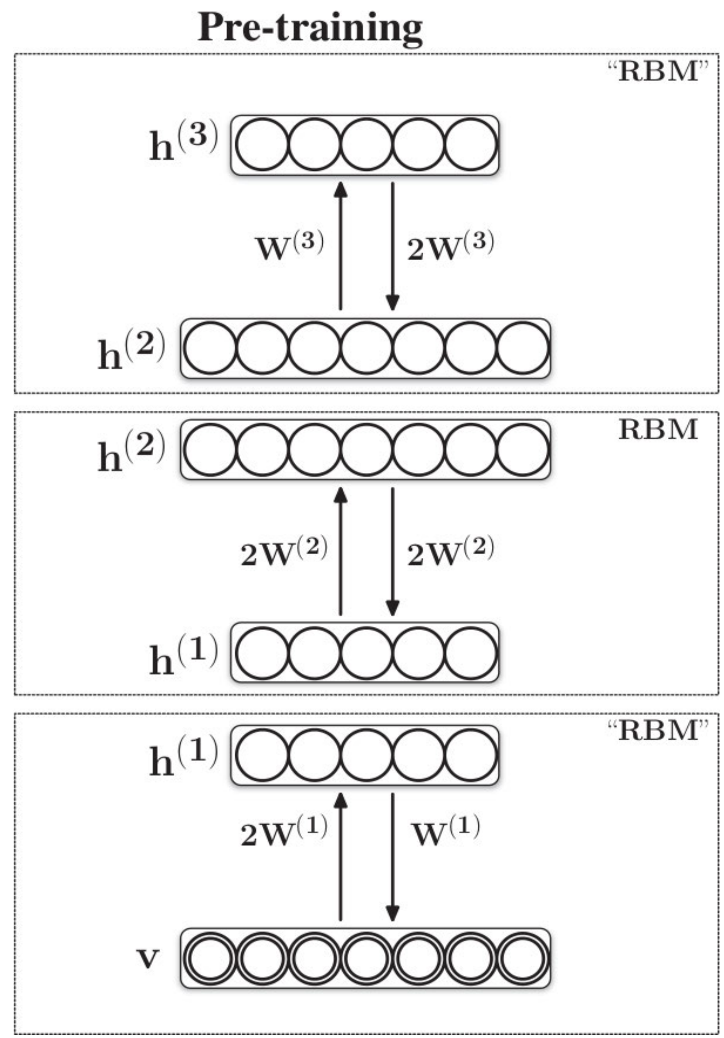
Recall: Error Backpropagation

- Used to learn a multi-layer perceptron
  - Biologically implausible for visual system:
    - „output units“ and error on these
    - back-propagation of error
  - Error back-propagates badly through many layers
- Internal representations emerge
  - but these are hard to interprete



Deep Learning





Learning of deep architectures:

1. unsupervised learning of individual hidden layers, layer by layer – guided by findings from biology
2. final supervised learning of the whole

Salakhutdinov, Hinton. An Efficient Learning Procedure for Deep Boltzmann Machines. MIT Tech Rep, 2010

Generative Models for Unsupervised Learning

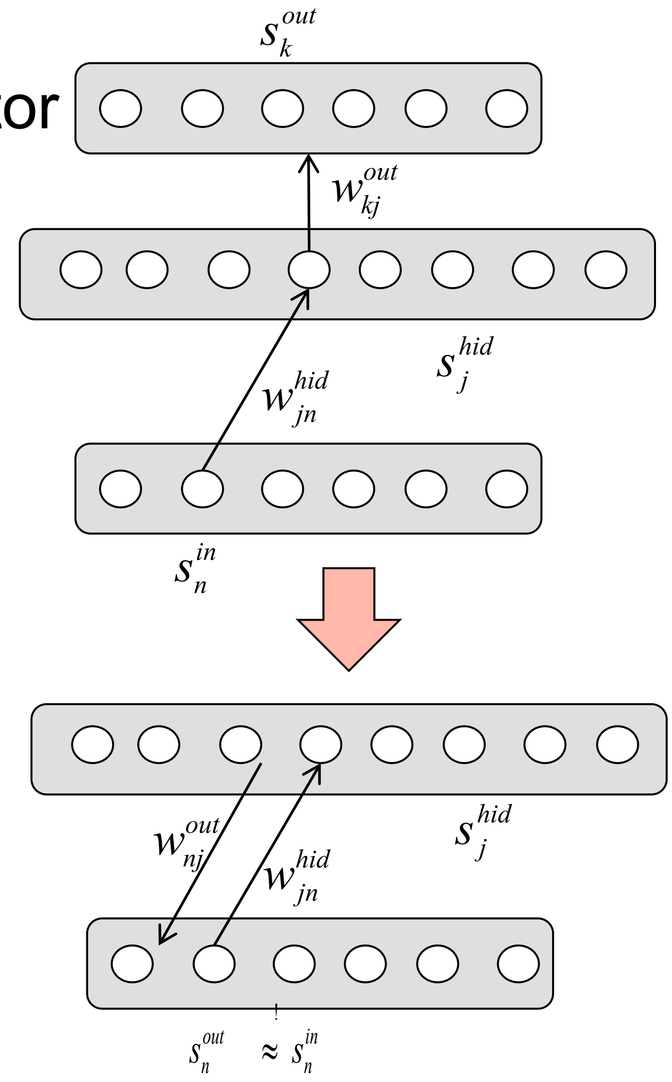
- Special case of MLP: auto-associator

$$s^{out} \overset{!}{\approx} s^{in}$$

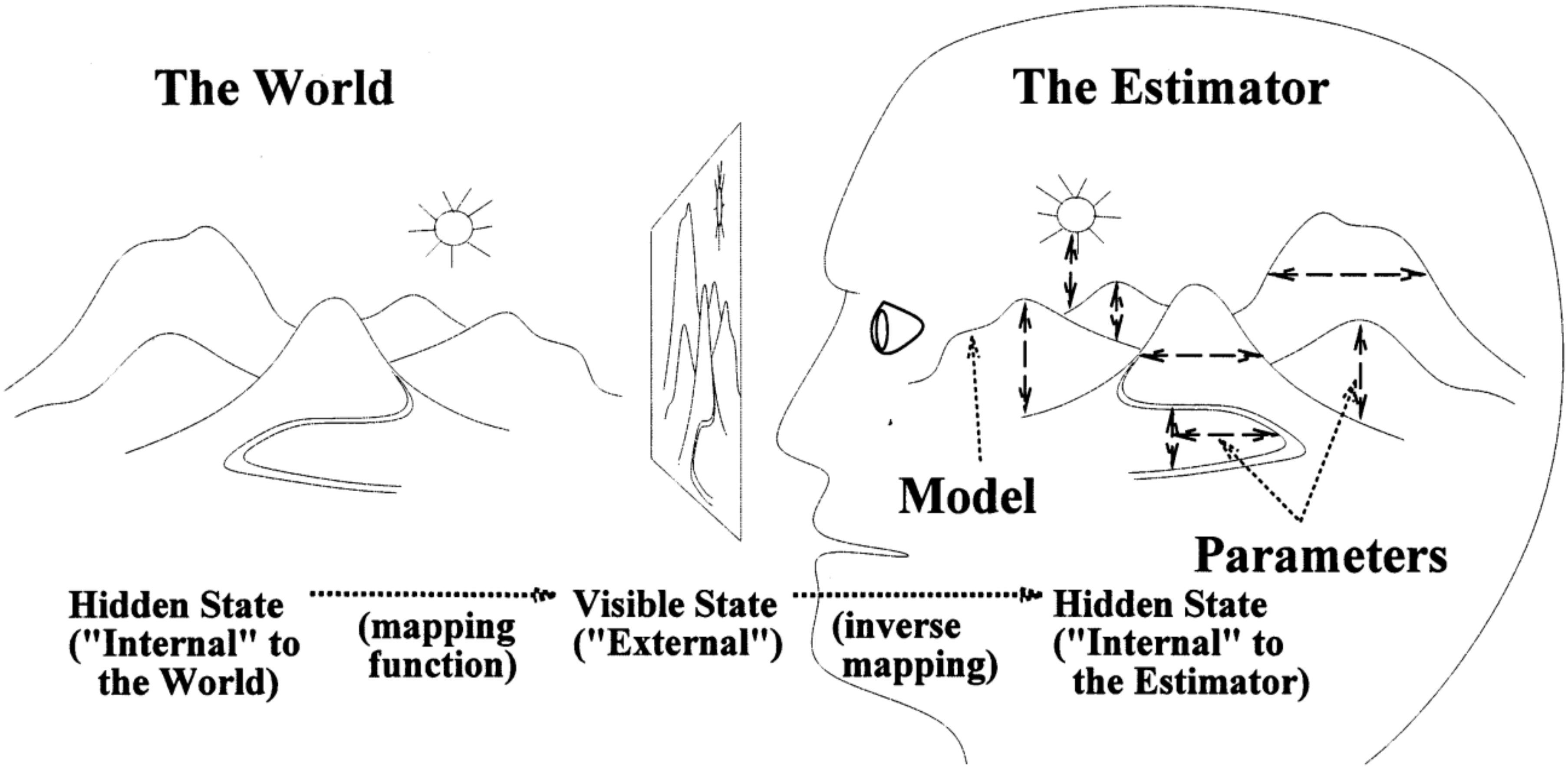
- We can now set:

$$W^{hid} = (W^{out})^T$$

  - (even though biologically unrealistic)
  - Avoids back-propagating the error through the layers
  - Train only one layer
  - Weights must scale to get the reconstruction right

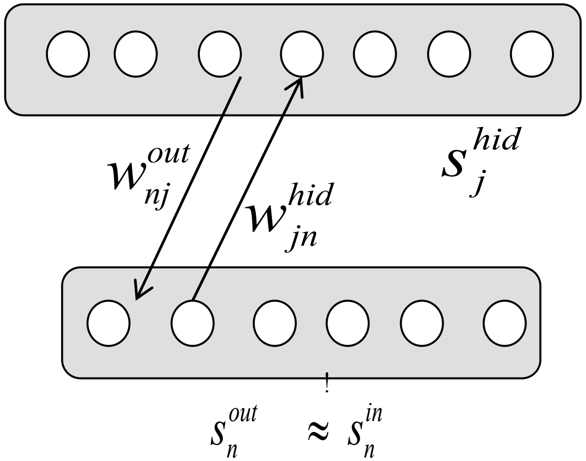


Generative Models



Rao. An optimal estimation approach to visual perception and learning. Vis Res, 1999

Generative Models



- Assume: $W^{hid} = (W^{out})^T$

  $\vec{s}^{\,hid} = g(W^{hid}\vec{s}^{\,in})$, $\quad \vec{s}^{\,out} = W^{out}\vec{s}^{\,hid}$

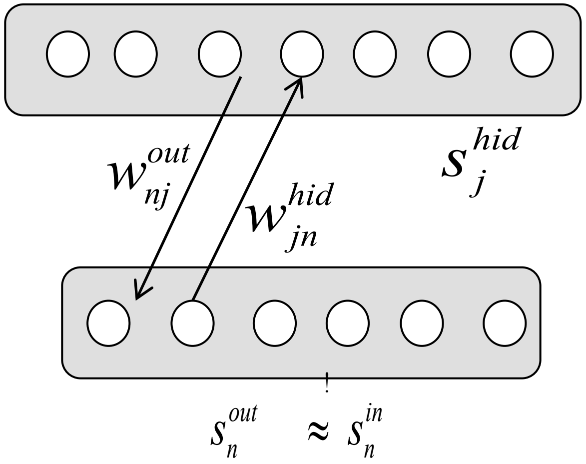
- Error function: $E(W, \vec{s}^{\,hid}) = \sum^{data} \frac{1}{2}(\underbrace{\vec{s}^{\,in} - \vec{s}^{\,out}}_{\vec{e}})^2$

- Activation update („E-step“): $\Delta\vec{s}^{\,hid} \approx -\frac{\partial E}{\partial \vec{s}^{\,hid}} = W^{hid}\vec{e}$
  - optimise for each data point

- Learning („M-step“): $\Delta W^{out} \approx -\frac{\partial E}{\partial W^{out}} = \vec{e}\vec{s}^{\,hid}$
  - slightly modify for one data point, using the optimal $\vec{s}^{\,hid}$

- Interpretation: $\vec{s}^{\,out} = W^{out}\vec{s}^{\,hid} = \sum_j^{n\_hid} \underbrace{\vec{w}_j^{\,out}}_{\text{basis functions}} s_j^{hid}$

Generative Models



- So far, we have set $w^{hid} = (w^{out})^T$
- Alternative: Wake-sleep algorithm.
- Wake phase learning step (similar to the previous):

$$\Delta w^{out} \approx e s^{hid}$$

- The sleep phase turns the model upside down:
  - Generate random activities $\tilde{s}^{hid}$
  - From these, generate „imagined" inputs: $\tilde{s}^{in} = W^{out} \tilde{s}^{hid}$
  - Sleep phase learning step:

$$\Delta W^{hid} \approx (\tilde{s}^{hid} - W^{hid} \tilde{s}^{in}) \tilde{s}^{in}$$

Hinton Dayan Frey Neal. The wake-sleep algorithm for unsupervised neural networks. 1995.
Weber. Self-organization of orientation maps, ... in the primary visual cortex. ICANN, 2001

Probabilistic Interpretation

- Likelihood of the data point $\vec{s}^{\,in}$ being generated, given model parameters $W, \vec{s}^{\,hid}, \ldots$ is:

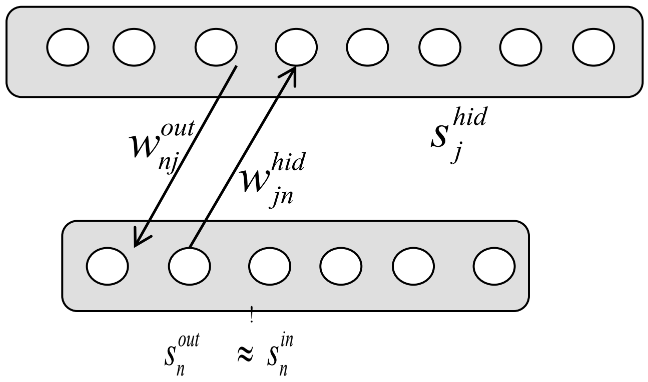
$$P(\vec{s}^{\,in} \mid W, \vec{s}^{\,hid}) \approx e^{-\frac{1}{2}(\vec{s}^{\,in} - \vec{s}^{\,out})^2} = e^{-E(W, \vec{s}^{\,hid})}$$

  - E=-ln(P)
  - $p^a p^b = e^{-E^a} e^{-E^b} = e^{-E^a - E^b}$
  - Multiplying probabilities $\Leftrightarrow$ adding cost terms
- we can define prior probabilities on parameters W and $\vec{s}^{\,hid}$

$\Leftrightarrow$ implement their effect with additional cost terms

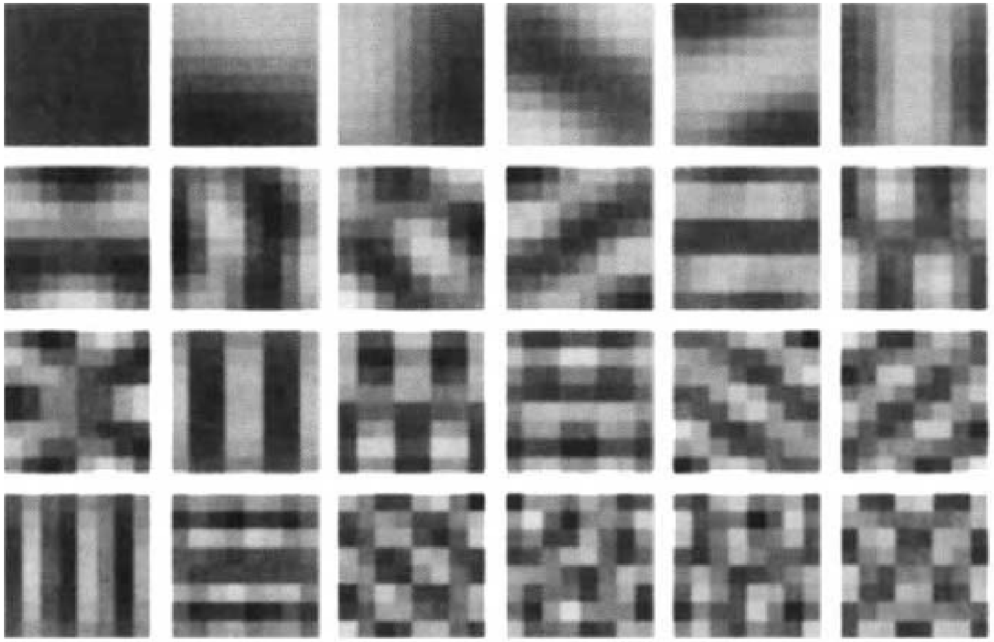
Generative Models

- additional cost terms / bottleneck lead to interesting coding on hidden area:
  - few hidden neurons – PCA
  - weight constraints
  - sparse hidden activations – ICA
  - non-negativity – NMF
  - denoising
  - winner code – vector quantization, k-means, Kohonen



Generative Models - PCA

- Bottleneck: few hidden neurons

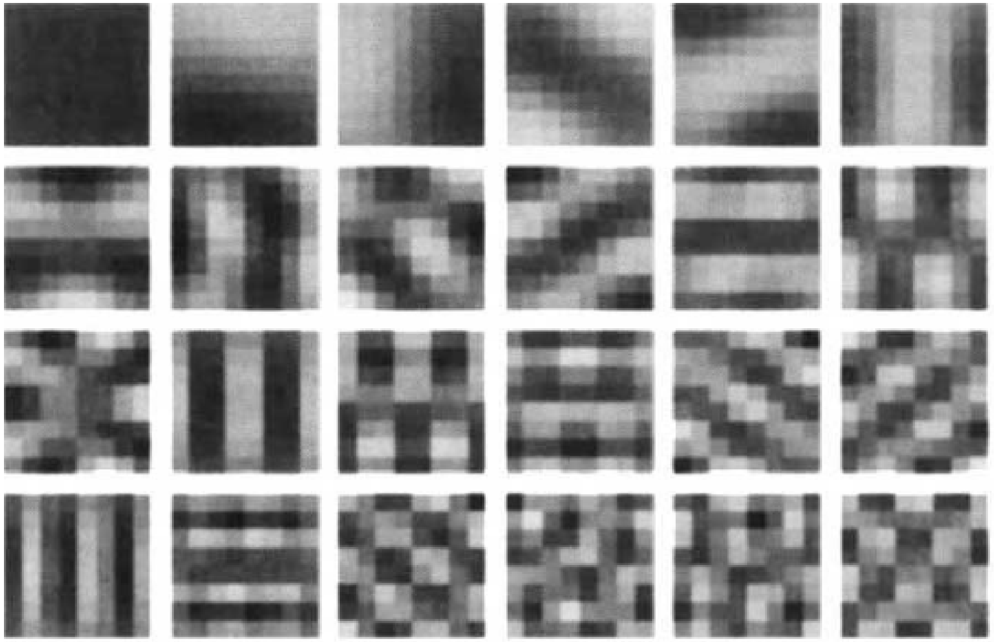


- Patches of natural images as training data
- Resulting basis functions resemble principle components

Olshausen, Field. Emergence of Simple-Cell Receptive Field Properties ... Natural Images. Nature, 1996

Generative Models - PCA

- Bottleneck: few hidden neurons



- Sanger's rule:

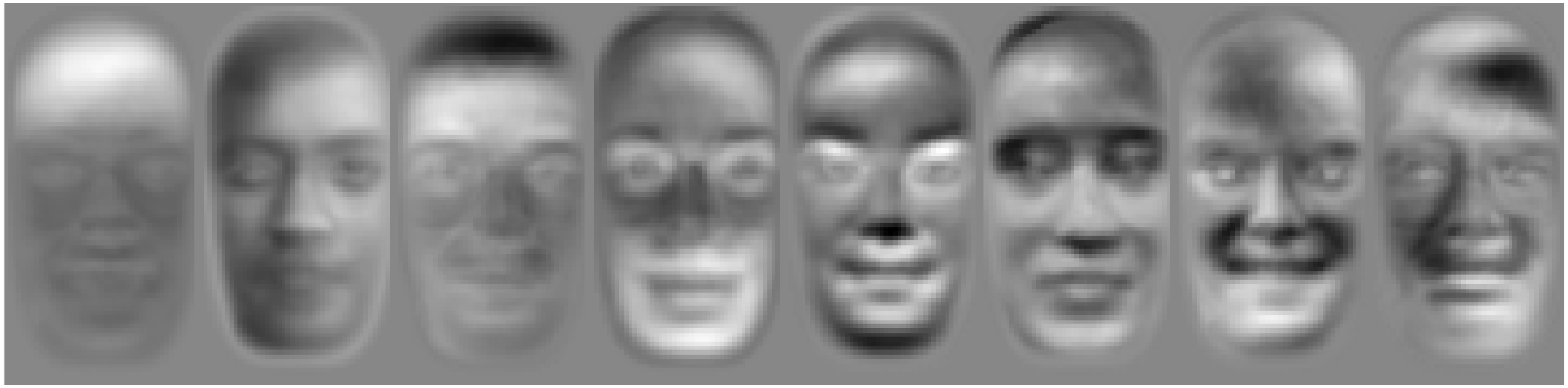
  (also sorts the neurons)

$$\Delta \vec{w}_i = \varepsilon(\vec{s}^{\,in} - \underbrace{\sum_{i'=1}^{i-1} s_{i'}^{hid} \vec{w}_{i'}}_{\text{feedback up to unit i-1}})s_i^{hid}$$

- E.g. unit 1: $\Delta \vec{w}_1 = \varepsilon \underbrace{\vec{s}^{\,in} s_1^{hid}}_{\text{Hebb}} = \varepsilon \vec{s}^{\,in} (\vec{w}_1^T \vec{s}^{\,in}) = \varepsilon \underbrace{\left\langle \vec{s}^{\,in} (\vec{s}^{\,in})^T \right\rangle \vec{w}_1}_{\text{correlation matrix}}$

Generative Models – PCA

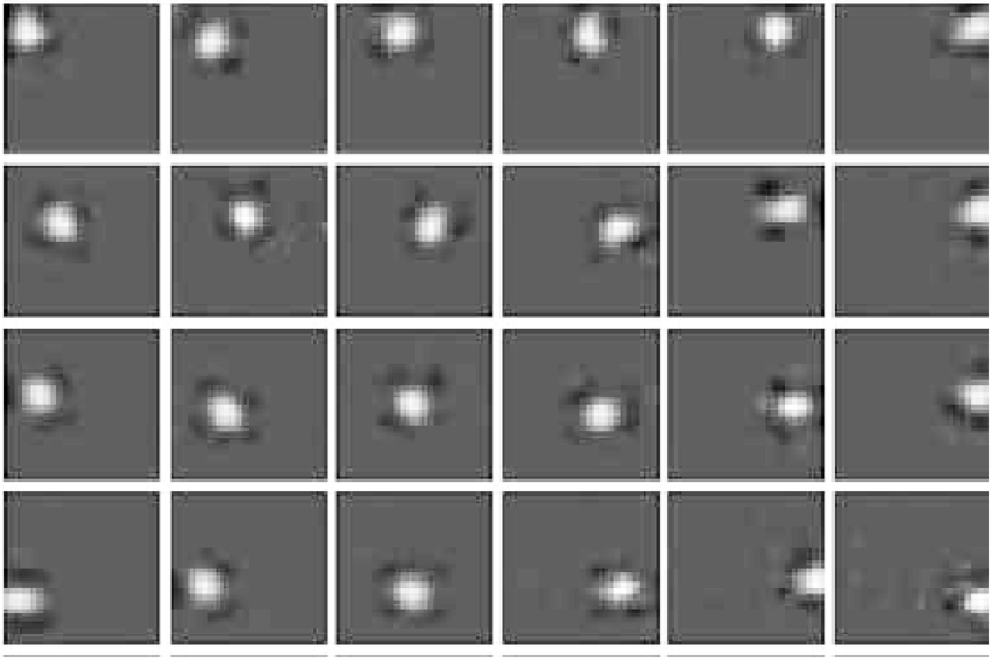
- Eigenfaces



http://students.washington.edu/ccgau/courses/CSE576/project4/ Jebara Pentland, Bayesian Face Recognition. Pattern Recog., 2000

Generative Models – Weight Limit

- Bottleneck: weight constraint (also penalizes large firing rate)

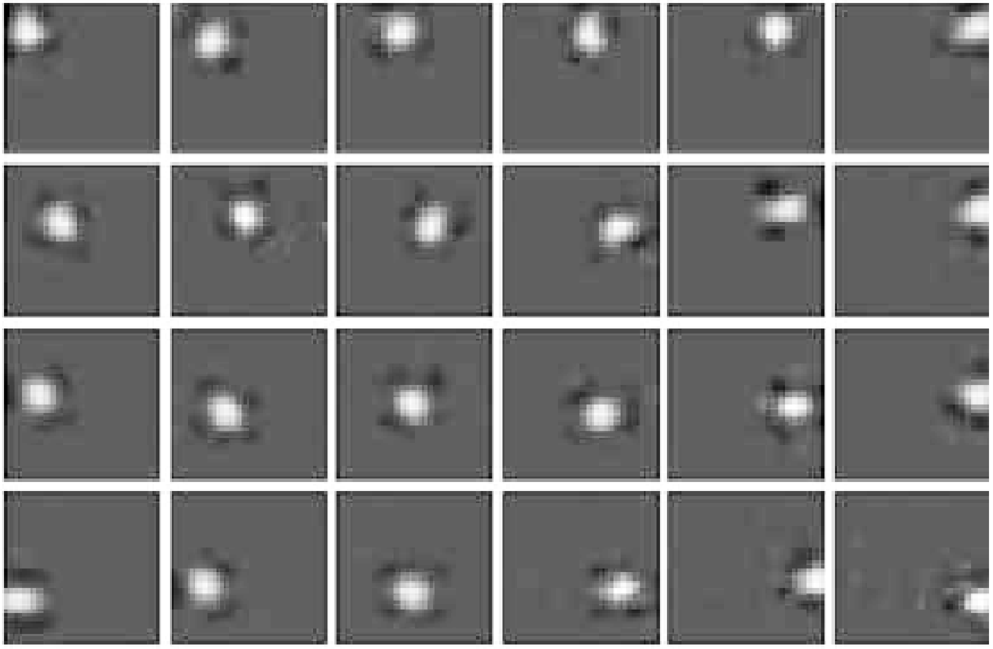


- natural images represented in centre-surround fashion
  → retinal ganglion cell receptive fields

Vincent, Baddeley. Synaptic energy efficiency in retinal processing. Vis. Res., 2003
Karklin, Simoncelli. Efficient coding of natural images with ... noisy Linear-Nonlinear neurons. NIPS, 2011

Generative Models – Weight Limit

- Implementation of the weight constraint



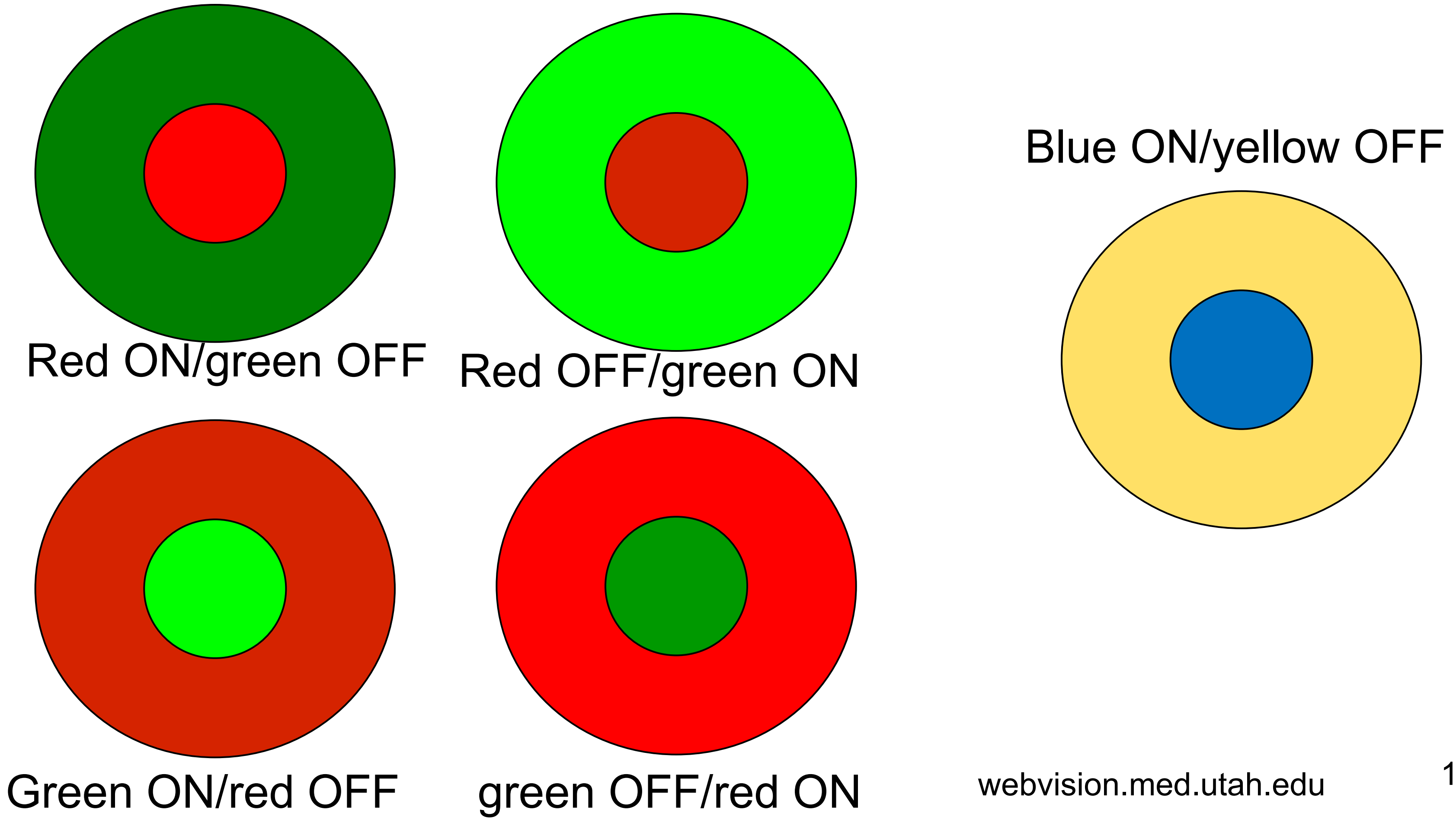
$$\Delta w_{ij} \approx \underbrace{e_j s_i^{hid}}_{good\ \ \text{reconstruction}} - \underbrace{c_1 * sign(w_{ij})}_{decay\ \ applied\ \ if\ \|\vec{w}_i\|_2 > c_2}$$

$$Note: -\frac{\partial}{\partial w} w^2 = -w \ldots \text{ term would penalize squared error on weights}$$

Vincent, Baddeley. Synaptic energy efficiency in retinal processing. Vis. Res., 2003

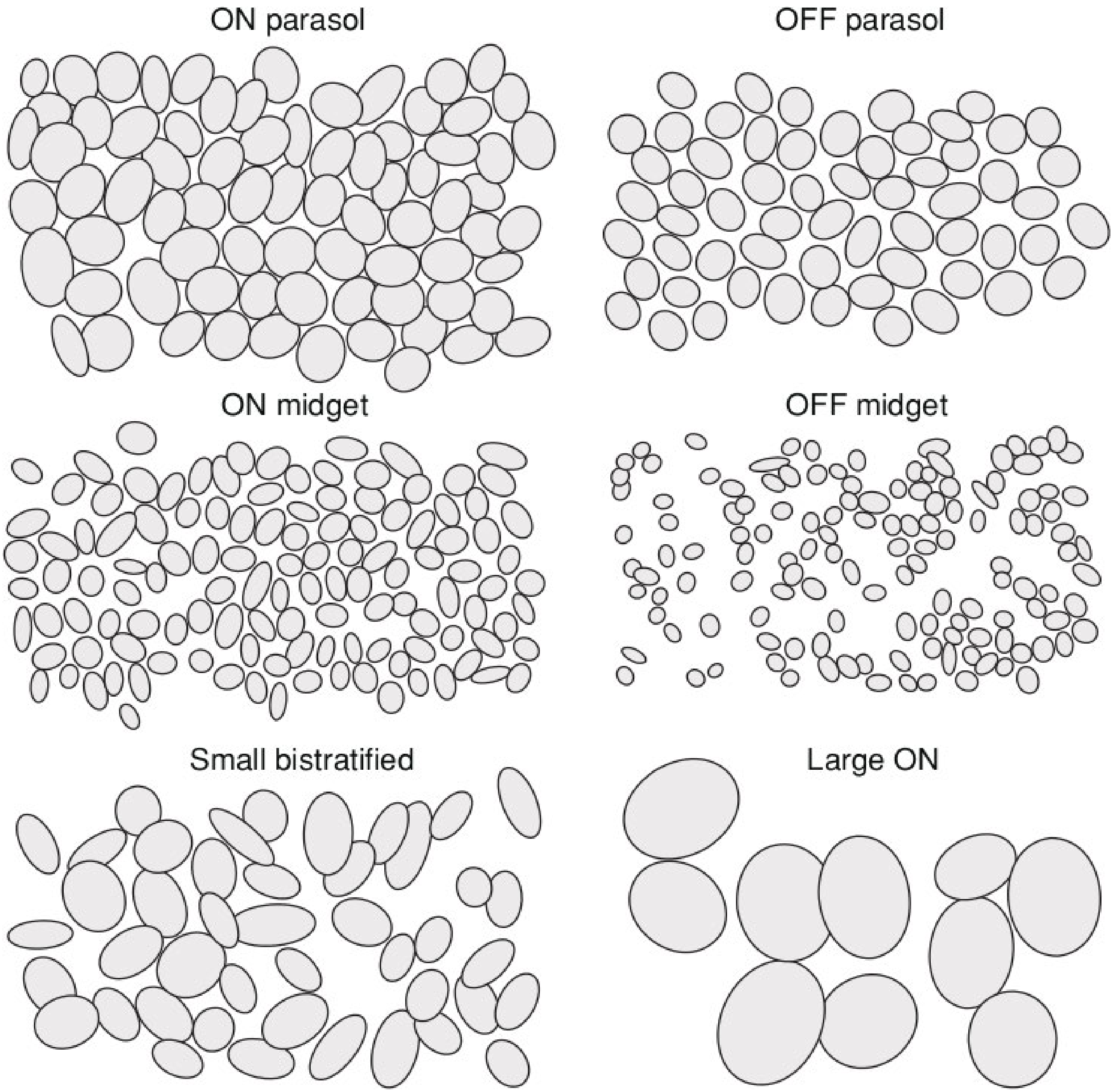
Generative Models – Weight Limit

- Possible results for color images:
  color-opponent ganglion cell receptive fields



webvision.med.utah.edu

Generative Models – Weight Limit

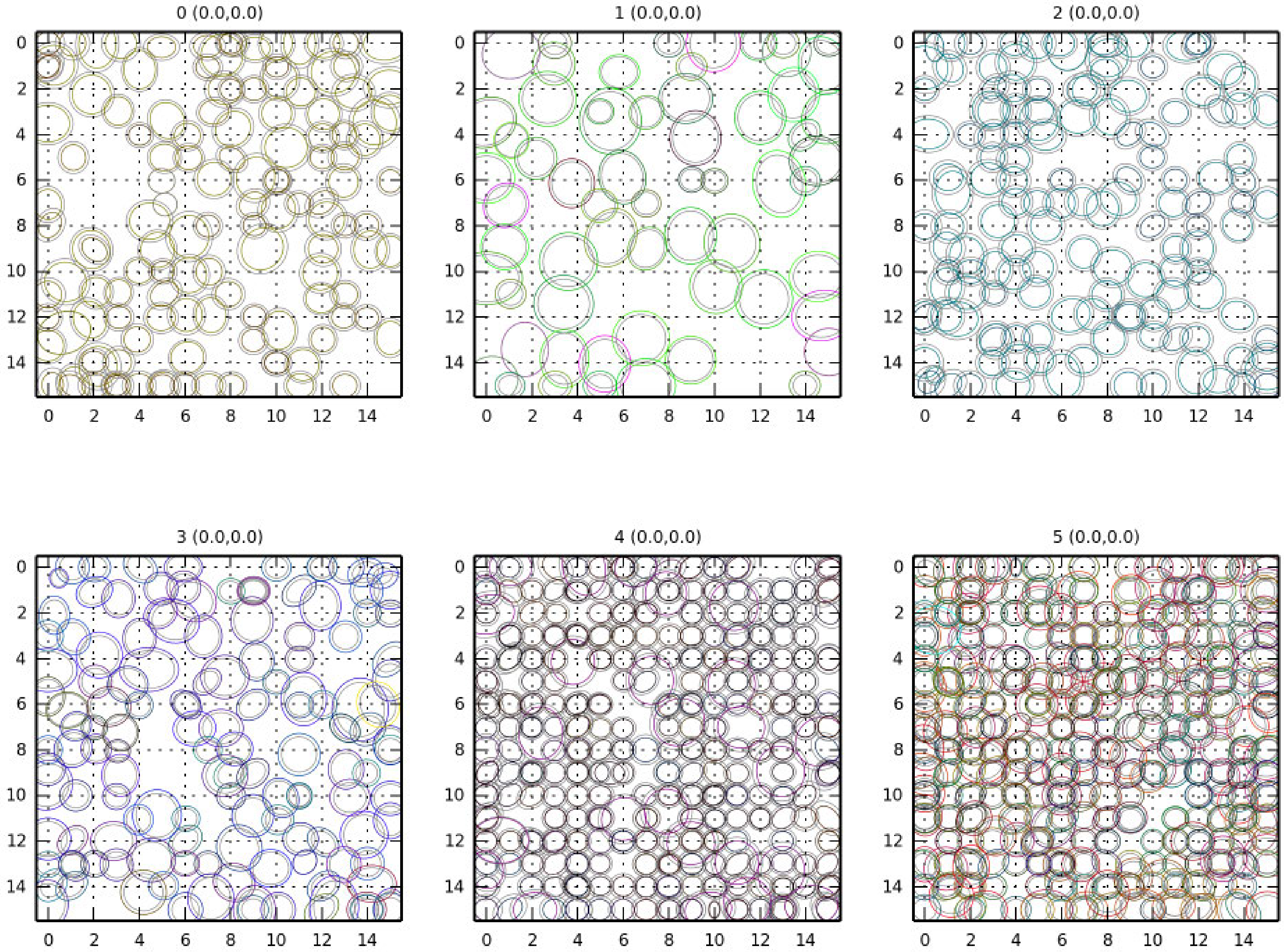


luminosity,
phasic

red-green

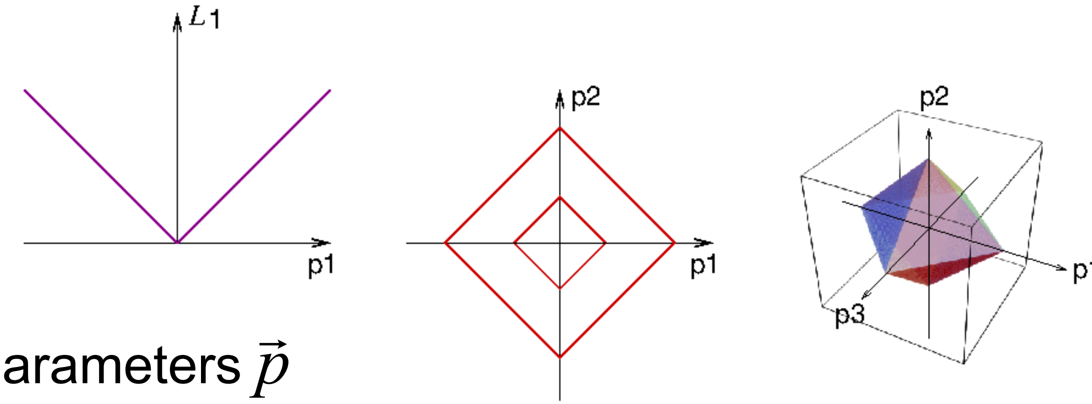
left:
blue-yellow

Generative Models – Weight Limit



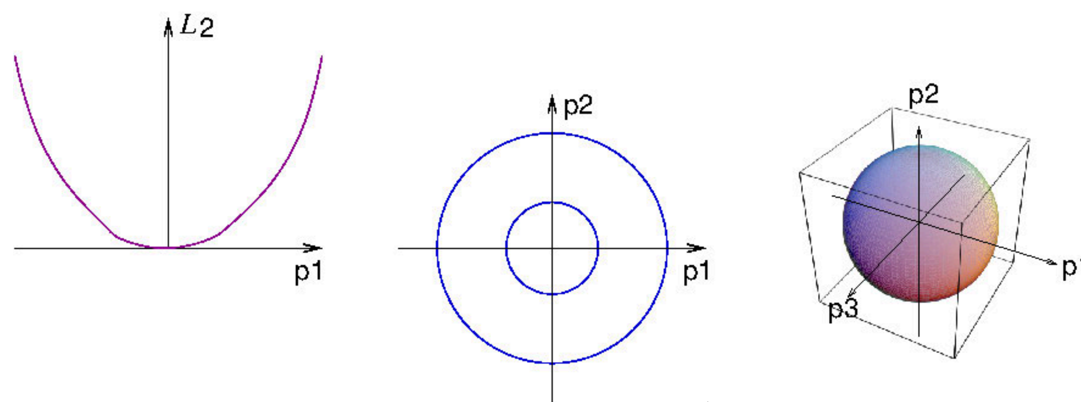
Constraints with L1 or L2 Norm

$$\| \vec{p} \|_1 = \sum_i | p_i |$$



→ L1 norm favors sparse parameters $\vec{p}$

$$\| \vec{p} \|_2 = \sqrt[2]{\sum_i p_i^2}$$

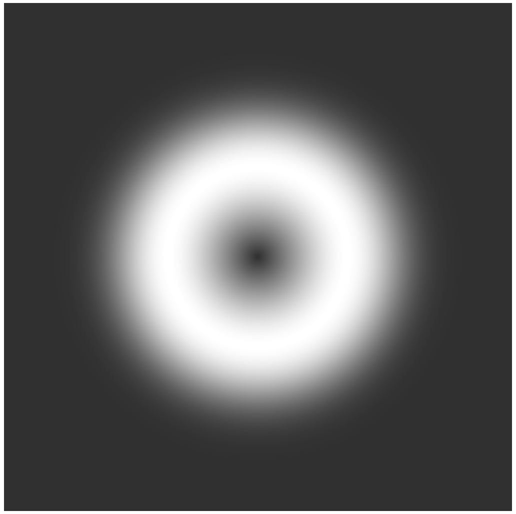


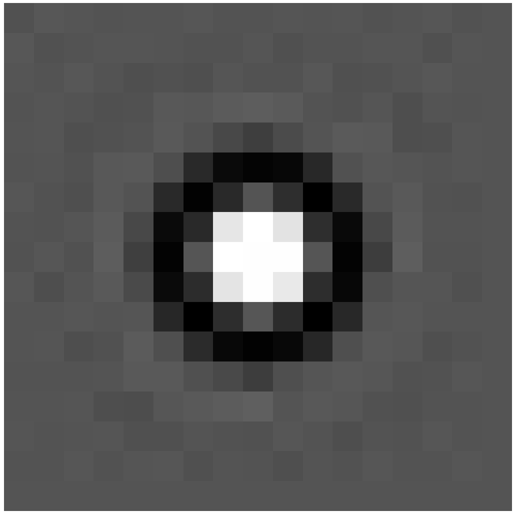
→ L2 norm penalises large parameters, but will not turn $\vec{p}$

$\| \vec{p} \|_\infty = \max_i \{| p_i |\} \rightarrow L\infty$ norm favours all p_i to be the same

Generative Models – Retinal Preprocessing

- pre-processing of input images by filtering spatial frequencies f
- Filter $R(f) = f \cdot e^{-(f/f_0)^4}$ has two terms:
  - f reduces low frequencies (equalizes the image amplitude spectrum of ≈ $1/f$)
  - $e^{-(f/f_0)^4}$ reduces high frequencies (pixel noise)

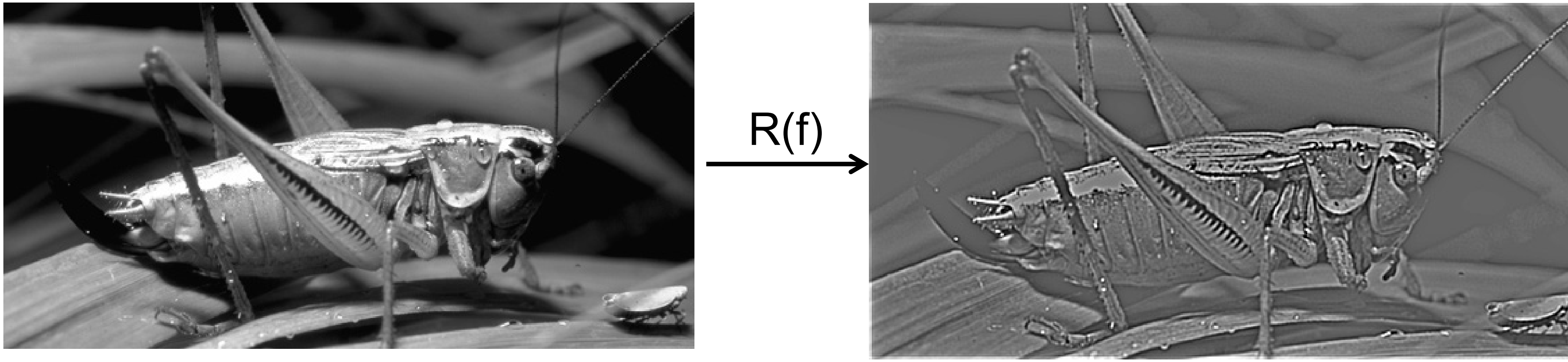




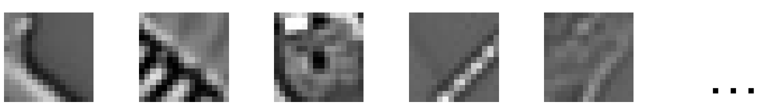
  - Filter in space ≈ retinal ganglion cell receptive fields

Olshausen, Field. Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1? Vis. Res., 1997

Generative Models – Retinal Preprocessing



- Next, a random part (to match the input layer size) is cut out:

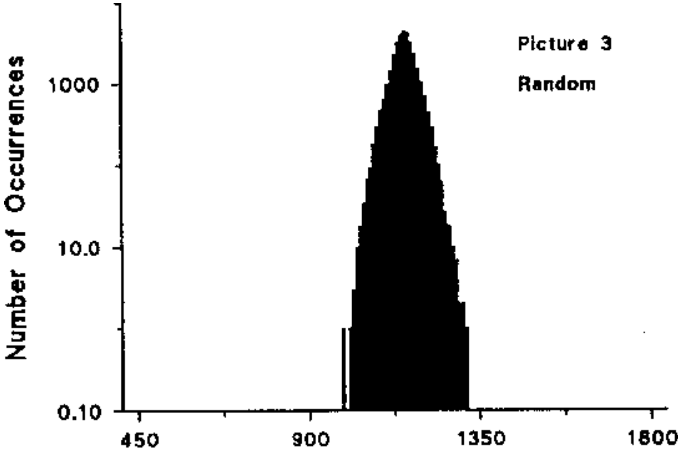


Olshausen, Field. Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1? Vis. Res., 1997

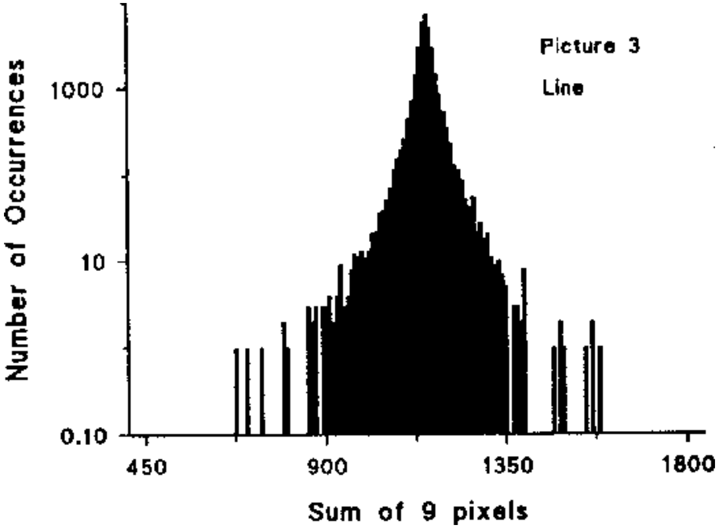
Generative Models – Sparse Coding

Centre-surround filtered image

Statistic of random patterns

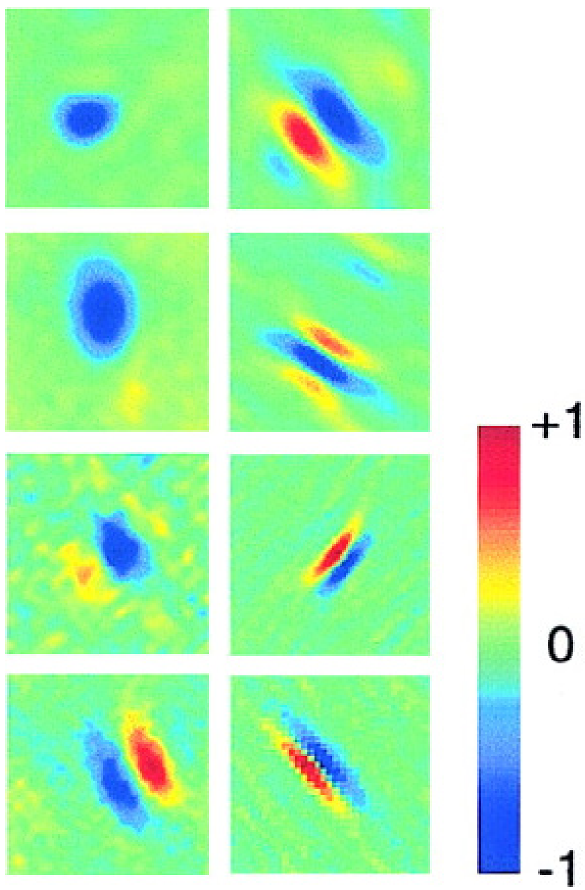


Statistic of edges

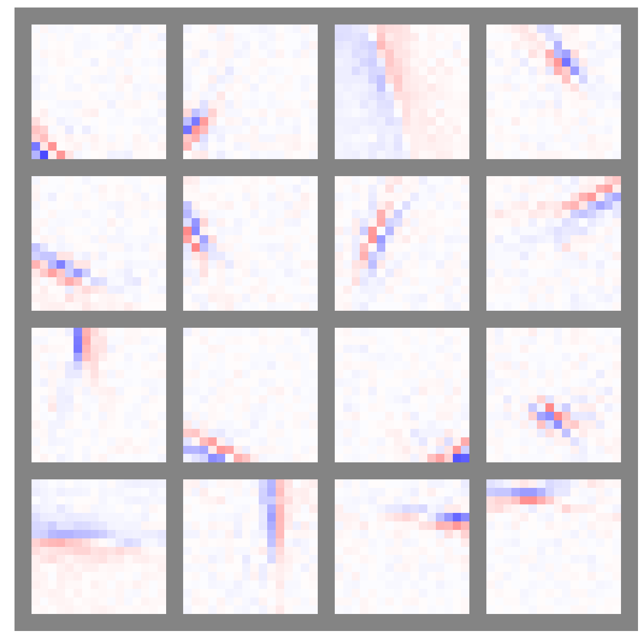


Barlow. Large Scale Neuronal Theories of the Brain., ch. What is the computational Goal of the Neocortex?, 1994

Generative Models – Sparse Coding



bottleneck:
sparse coding

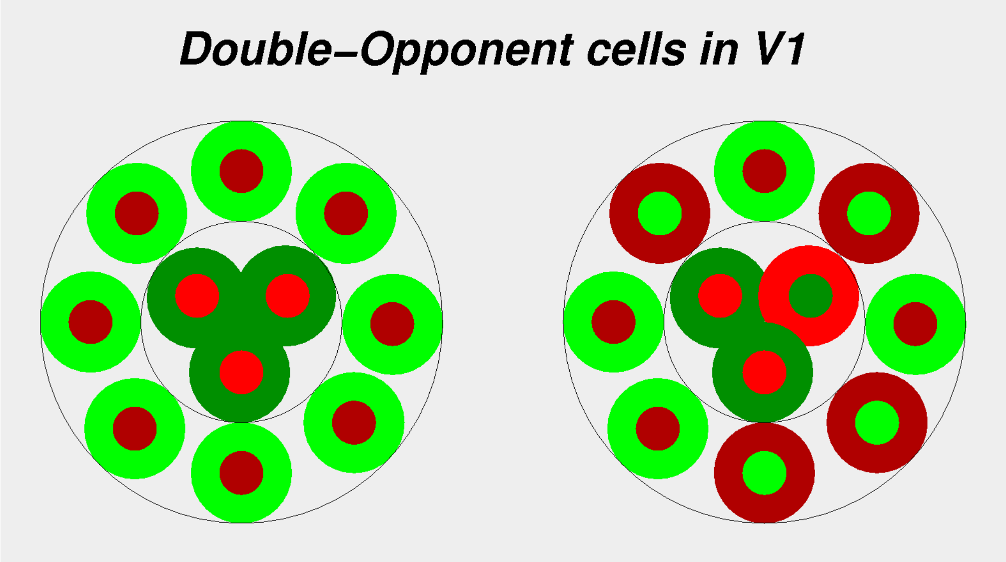


pre-processed natural images represented as localized edges
→ primary visual cortex (V1) cell receptive fields

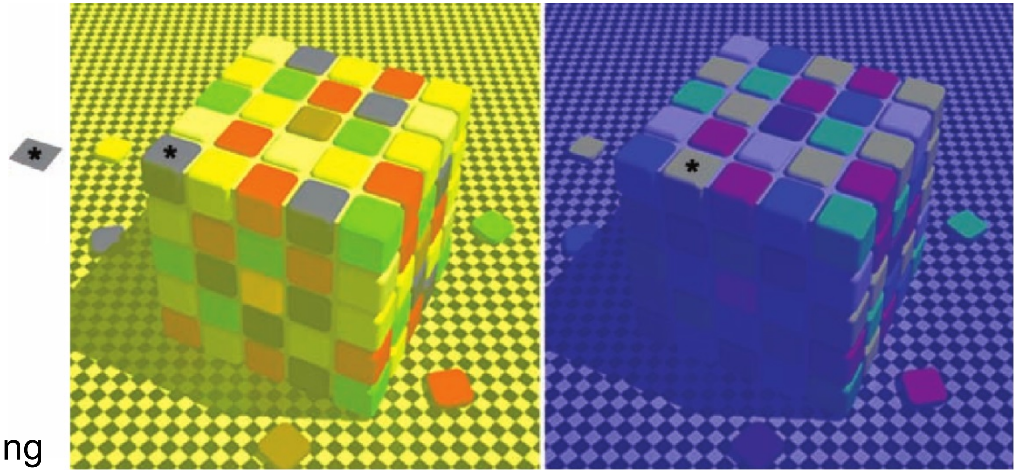
Ringach. Spatial Structure ... of Simple-Cell Receptive Fields in Macaque Primary Visual Cortex. J Neurophysiol. 2002 Olshausen, Field. Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1? Vis. Res.,1997

Generative Models – Sparse Coding

possible results for color images



?
⇒ color constancy

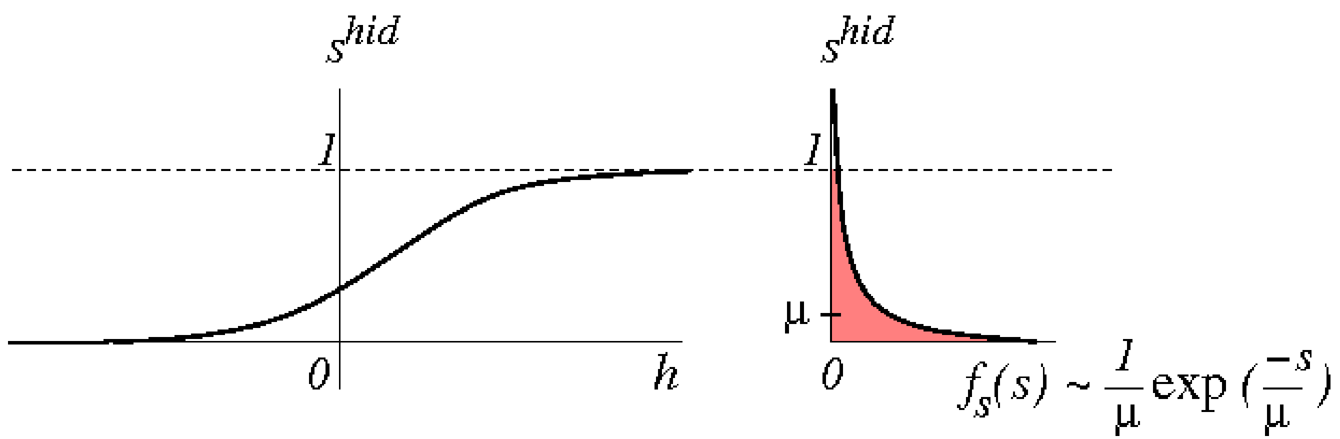


Conway. Spatial Structure of Cone Inputs to Color Cells in Alert Macaque Primary Visual Cortex. J Neurosci, 2001
Conway. Color Vision, Cones, and Color-Coding in the Cortex. The Neuroscientist, 2009

Generative Models – Sparse Coding

- how to impose sparseness on the hidden units' activations?
- transfer function has parameters, "slope" *a* and "threshold" *b*:

$$s_i^{hid} = g_{a,b}(h_i) = \frac{1}{1+e^{-(a_i h_i + b_i)}}$$



- set these intrinsic neuron's parameters such that the activity distribution $f(s_i^{hid})$ is close to an exponential with small mean μ

Weber, Triesch. A Sparse Generative Model of V1 Simple Cells with Intrinsic Plasticity. Neur Comp, 2008

Generative Models – Sparse Coding

- method 1: adapt the parameters a_i and b_i on-line:

$$\Delta a_i = \eta_a\left(\frac{1}{a_i} + h_i - 2h_i s_i - \frac{1}{\mu} h_i s_i + \frac{1}{\mu} h_i s_i^2\right)$$

$$\Delta b_i = \eta_b\left(1 - 2s_i - \frac{1}{\mu} s_i + \frac{1}{\mu} s_i^2\right)$$

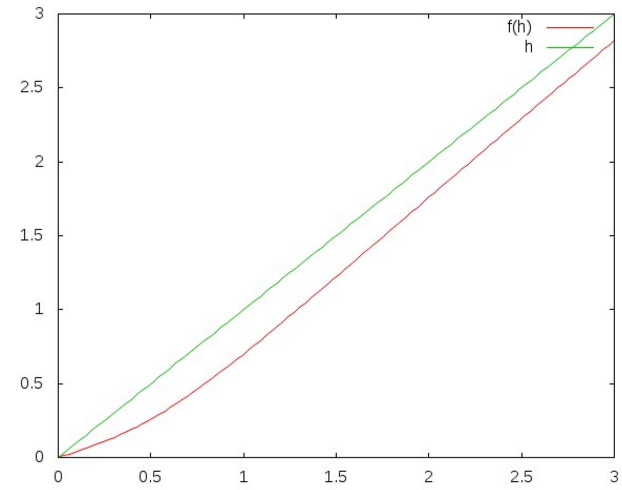
  - a_i and b_i converge much faster than the weights
  - $\vec{w}_i^{hid}$ must be normalized (because a_i effectively scales it)
  - separate weights W^{out} needed (e.g. use wake sleep algorithm)
  - different units will adopt individual parameters

Weber, Triesch. A Sparse Generative Model of V1 Simple Cells with Intrinsic Plasticity. Neur Comp, 2008

Generative Models – Sparse Coding

- method 2: adapt neuronal transfer function parameters "by hand" for sparseness

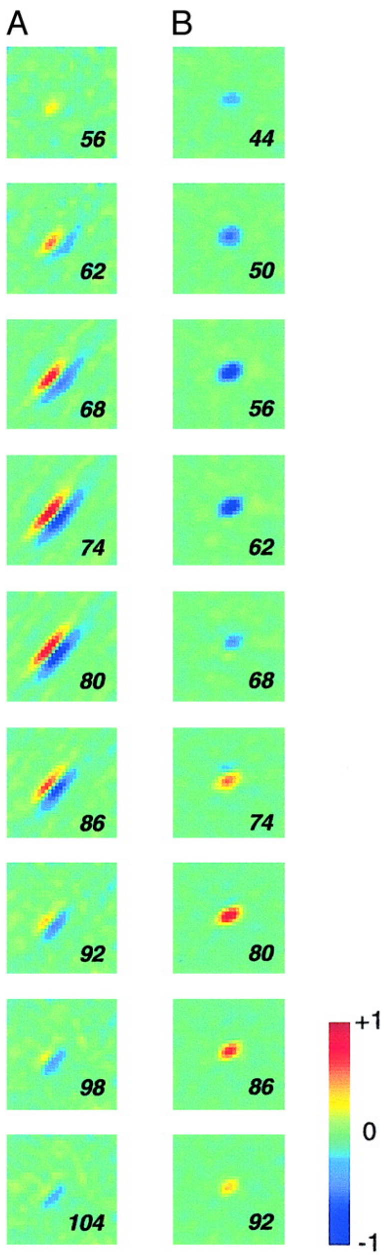
$$f(h) = h - \frac{0.3h}{1+h^2}$$



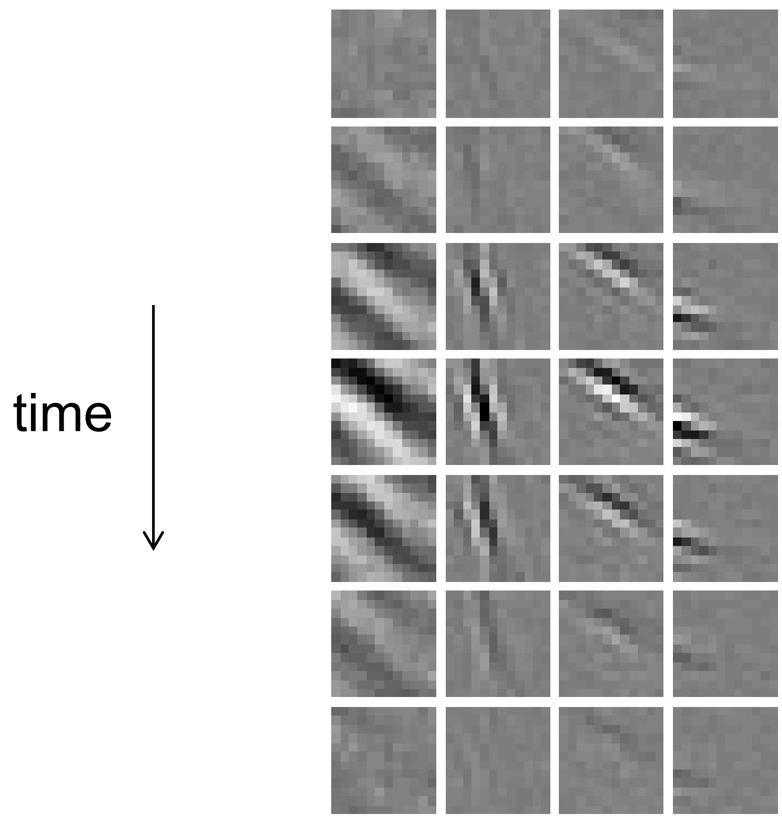
  - weight decay term needed (weights would compensate for sparseness constraint)
  - $w^{out} = (w^{hid})^T$

Olshausen, Field. Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1? Vis. Res., 1997

Sparse Coding



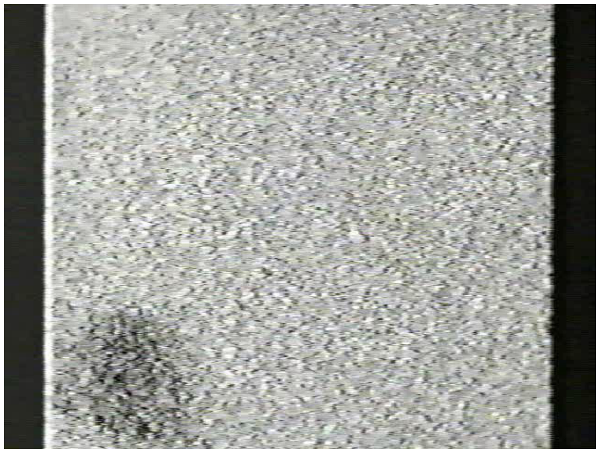
generative model for movies
→ spatiotemporal V1 receptive fields



Olshausen. Learning Sparse, Overcomplete Repr. of Time-Varying Natural Images. Conf Image Processing. 2003
Ringach. Spatial Structure ... of Simple-Cell Receptive Fields in Macaque Primary Vis. Cort. J Neurophysiol. 2002

Generative Models – Sparse Coding

biological training data:
pre-natal retinal waves



→ image-like properties:
- topographical relations
- edges
- convex figures
- coherent motion

...

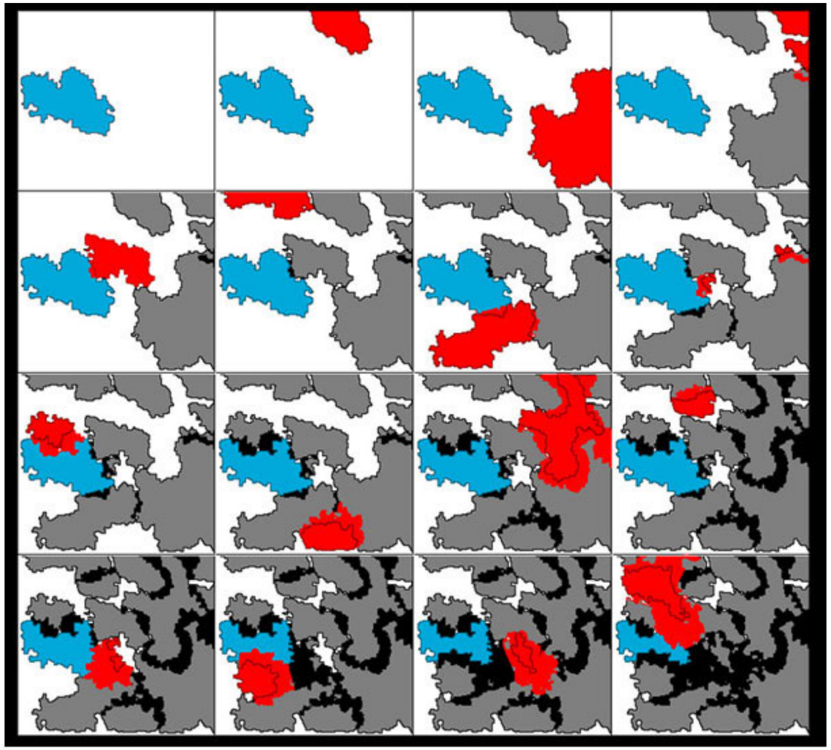


Fig. 5 . Spatiotemporal properties of retinal waves. Pictured is a sequence of domains that were measured in a single imaging field of view (read from left to right, top row first). The red domain in each frame corresponds to a new wave in the region. Overlapping regions in which more than one wave occurs are shown in black. The first wave in th e sequence is shown in blue. The entire sequence corresponds to 90 seconds of recording, and the total field of view is 1.2 x 1.4 mm. The propagation boundaries of waves are determined in part by wave-induced refractory regions that last for 40-50 seconds. Wave initiation sites as well as the locations of the boundaries are non-repeating (Feller et al, 1997).

note: "In the mature retina, the dendrites of On and Off ganglion cells are segregated into separate sublaminas of the inner plexiform layer, but early in development these processes are multistratified ... ganglion cells with multistratified dendrites respond to the onset, as well as the offset, of light."
pic: webvision.umh.es/webvision/DEV2.html; video: Feller, Wellis, Stellwagen, Werblin, Shatz. Requirement for cholinergic synaptic transmission in the propagation of spontaneous retinal waves. Science, 1996; quote: Wang, Liets, Chalupa. Unique Functional Properties of On and Off Pathways in the Developing Mammalian Retina, JNeurosci 2001

NEWS

New Sony In-Utero TV To Entertain Children In The Womb

JANUARY 27, 1999 | ISSUE 35•03

LOS ANGELES—The entertainment industry is abuzz following the Sony Corporation's unveiling Monday of the Utertron 9000, a state-of-the-art in-utero womb-entertainment system for children between the ages of minus nine months and zero.

Enlarge Image

The new Sony Utertron 9000, a state-of-the-art television for the unborn.

The Utertron 9000, touted as the first prenatelevision technology ever developed, features micro-miniature multi-speaker SurrounSound, HDTV compatibility and a luxurious one and one-sixteenth-inch screen. It will make its official debut Feb. 15 at the World OB/GYN Electronics Expo in Las Vegas, and is slated to hit stores nationwide by early fall.

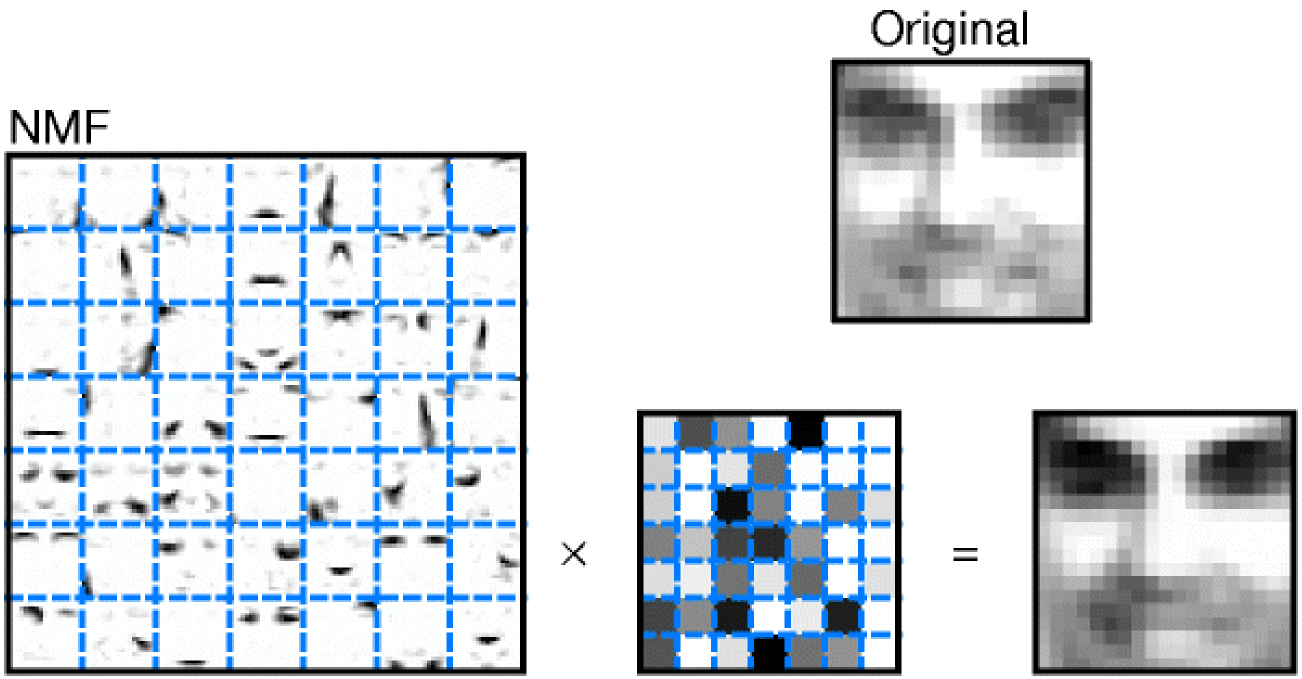
"Thanks to the revolutionary new Utertron 9000," Sony vice-president of media relations Grant Bellows told reporters, "developing fetuses everywhere will soon be able to enjoy the same high-quality entertainment programming that, until now, has only been available to the post-born."

Enlarge Image

An X-ray of a developing fetus enjoying a favorite television program on his or her Sony Utertron 9000.

Generative Models – NMF

bottleneck:
non-negativity
→ non-negative
matrix factorization

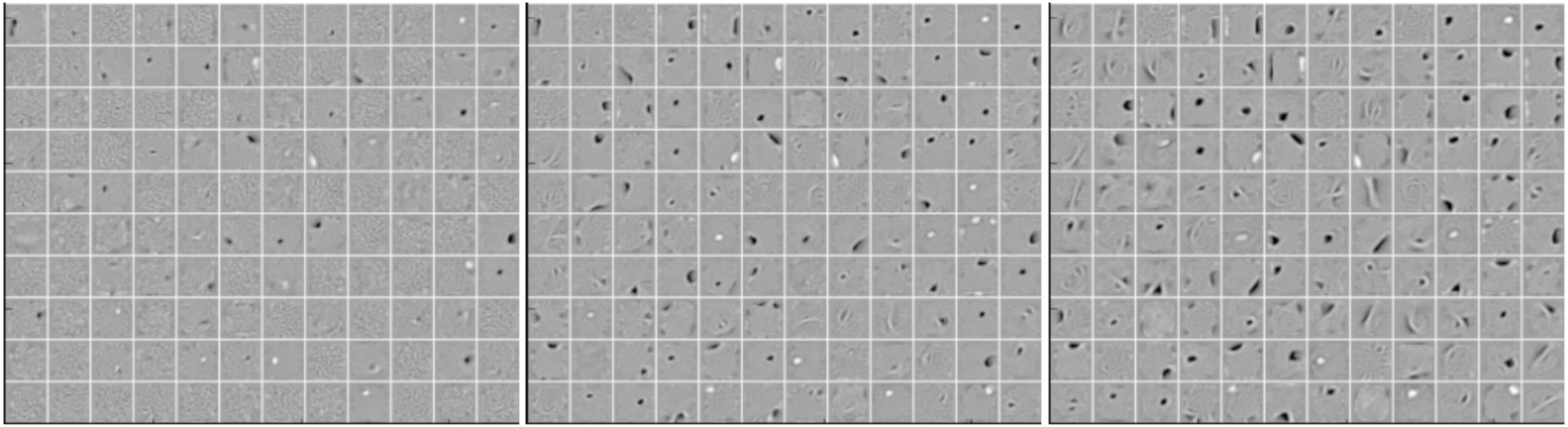


centered faces
→ part-based representations

Lee, Seung. Learning the parts of objects by non-negative matrix factorization. Nature, 1999

Generative Models – Denoising

- reconstruction from partially corrupted input patterns



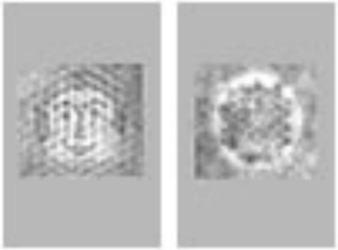
(a) No destroyed inputs (b) 25% destruction (c) 50% destruction

(input are MNIST digit data corrupted with zero-value “blank” pixels)

Vincent, Larochelle, Bengio, Manzagol. Extracting and Composing Robust Features with Denoising Autoencoders. Int Conf on Machine Learning, 2008

Newborn Looking Preferences to Faces

Table 1. *Number of babies who preferred each stimulus over its paired stimulus*



	Feature inversion			Phase and amplitude reversal			Contrast reversal		
Age	*Config*	Inversion	Neither	Phase of face	Amplitude of face	Neither	Positive contrast	Negative contrast	Neither
Newborns	9*	1	2	0	9*	3	0	0	12
6-week-olds	0	0	12	12**	0	0	3	0	9
12-week-olds	0	1	11	12**	0	0	12**	0	0

model: newborn’s face selective neuron has incoming weights arranged in a top-heavy triangle

Data: Mondloch et al. Face Perception During Early Infancy. Psychol. Science, 1999
Model: Bednar, Miikkulainen. Learning Innate Face Preferences. Neural Computation, 2003
However see: Wilkinson, Metta, Gredeb¨ack. Inter-facial relations: Binocular geometry when eyes meet. Proc ICMC Internatuional Conference on Morphology and Computation, 2011

Generative Models in Vision – Summary

- hierarchical visual system with growing abstraction
- weight matrices transform the representations
- generative model for learning
- prior constraints on hidden encoding:
  - few hidden neurons – PCA
  - weight constraints
  - sparse hidden activations – ICA
  - non-negativity – NMF
  - denoising
  - winner code – vector quantization, k-means, Kohonen‘s SOM

- innate face preferences
- peripheral / foveal preferences
- slow / fast responses