Neural Networks

Lecture 08

Neural Representations in the Visual System –

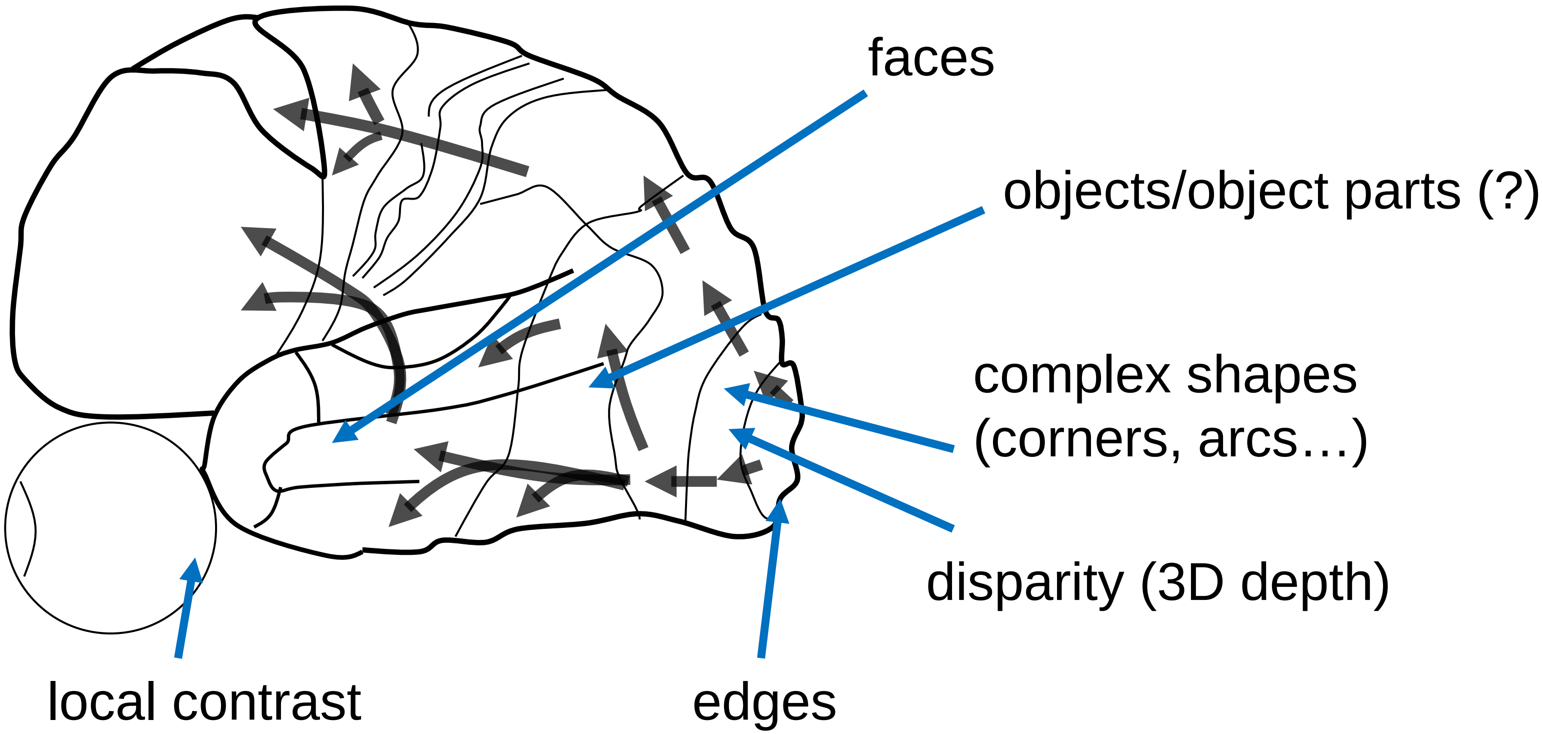
Unsupervised Learning with Generative Models

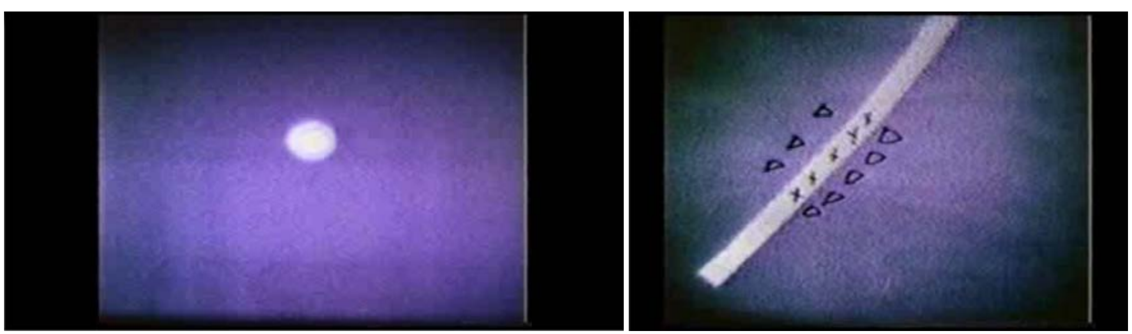
http://www.informatik.uni-hamburg.de/WTM/

Overview

- **Hierarchical visual system**
- **Generative auto-encoder architectures**
  - MLP backpropagation
  - Transposed weights model
  - Helmholtz machine
- **Constraints**
  - few hidden neurons
  - weight constraints
  - sparse hidden activations
  - non-negativity
  - denoising
  - other

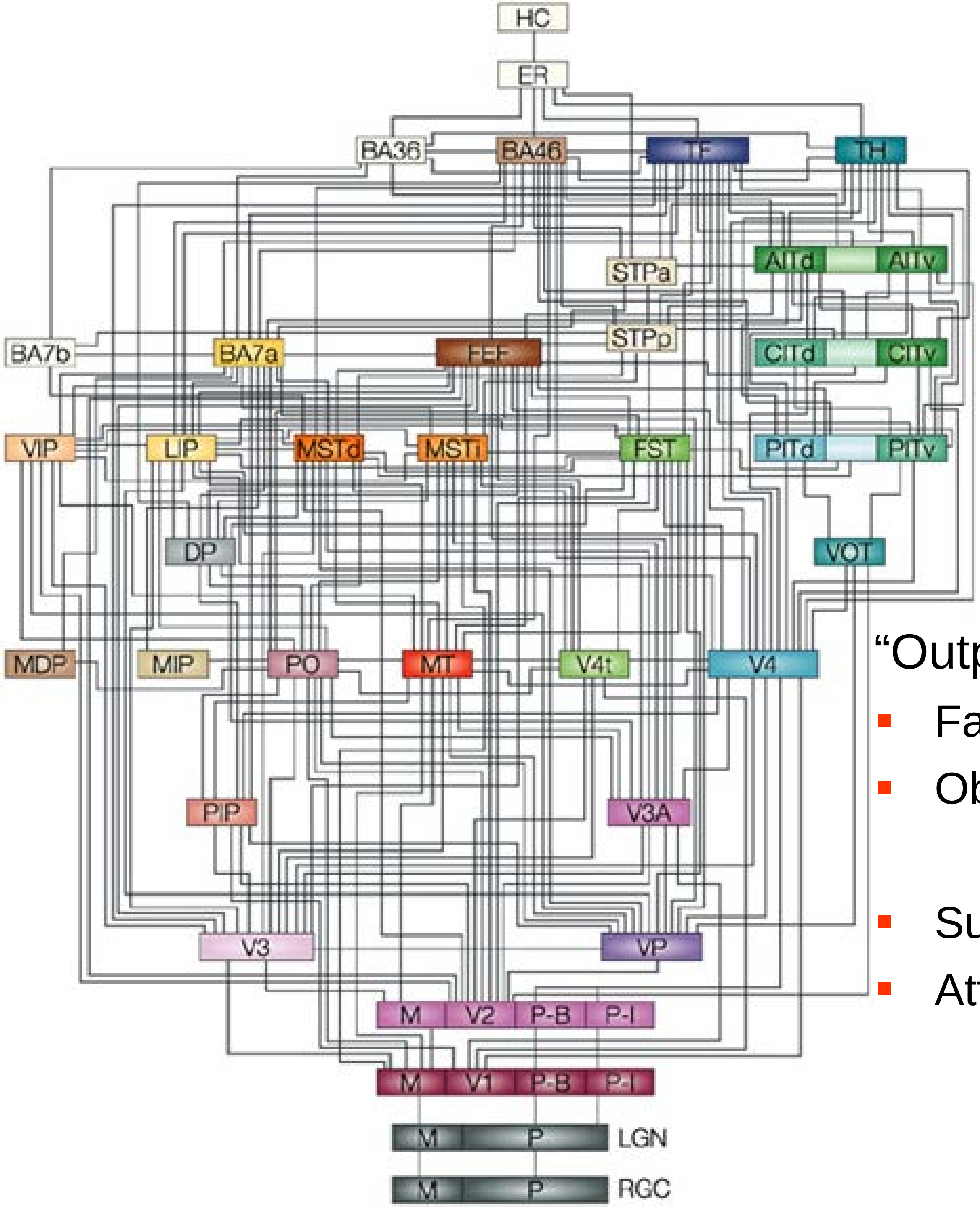
Representation in the Visual System of the Brain





Videos by Hubel & Wiesel

Hierarchy of the Visual System



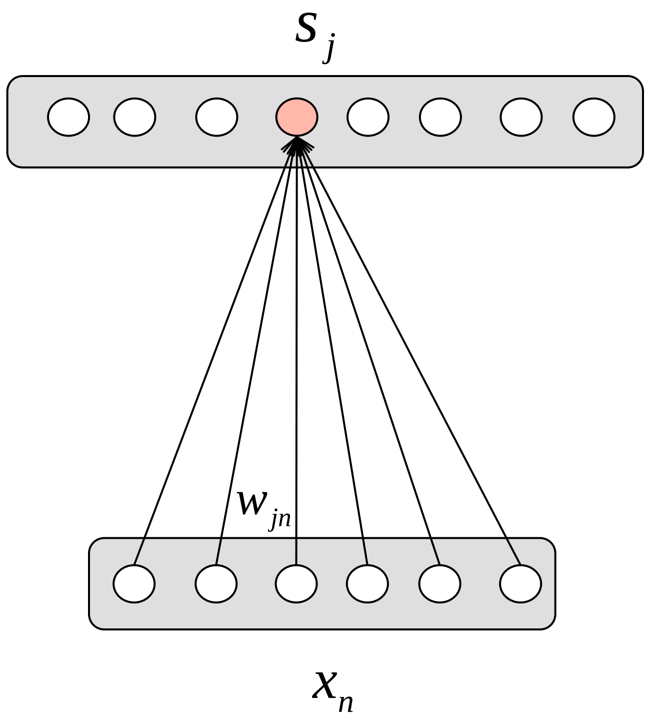
"Outputs" are diverse:

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

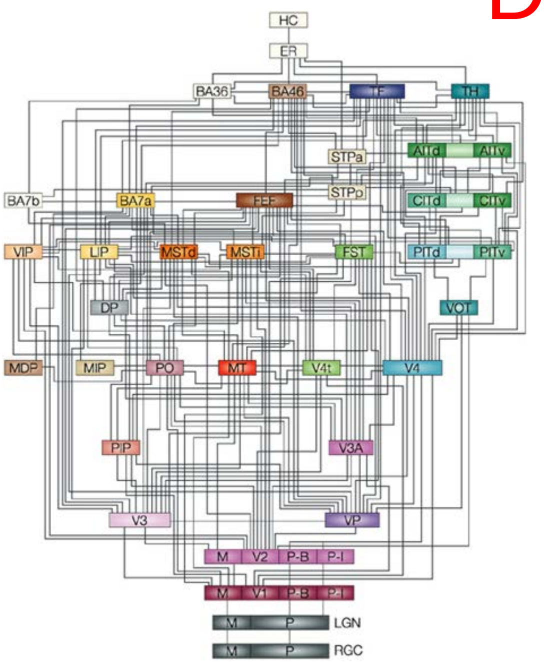
Recall: Perceptron/Connectionist Neurons

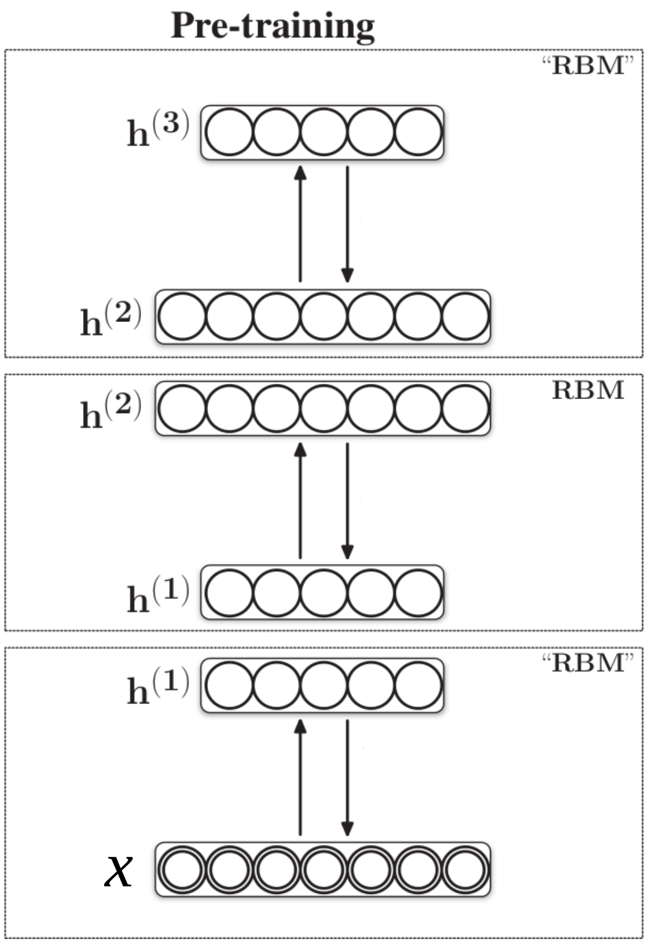
- Activate one neuron: $h_j = \sum_n w_{jn} x_n = \vec{w}_{jn} \cdot \vec{x}$
  - Dot product between weight vector and input vector
- Activate all neurons: $\vec{h} = W\vec{x}$
  - Matrix product with weight matrix
- In Python: h = numpy.dot(W,x)
- In C: two nested for-loops
  - Outer loop over output neurons, inner loop does scalar product
- Then transfer function applied, e.g.:

$$s_j = \tanh(h_j)$$



Deep Learning



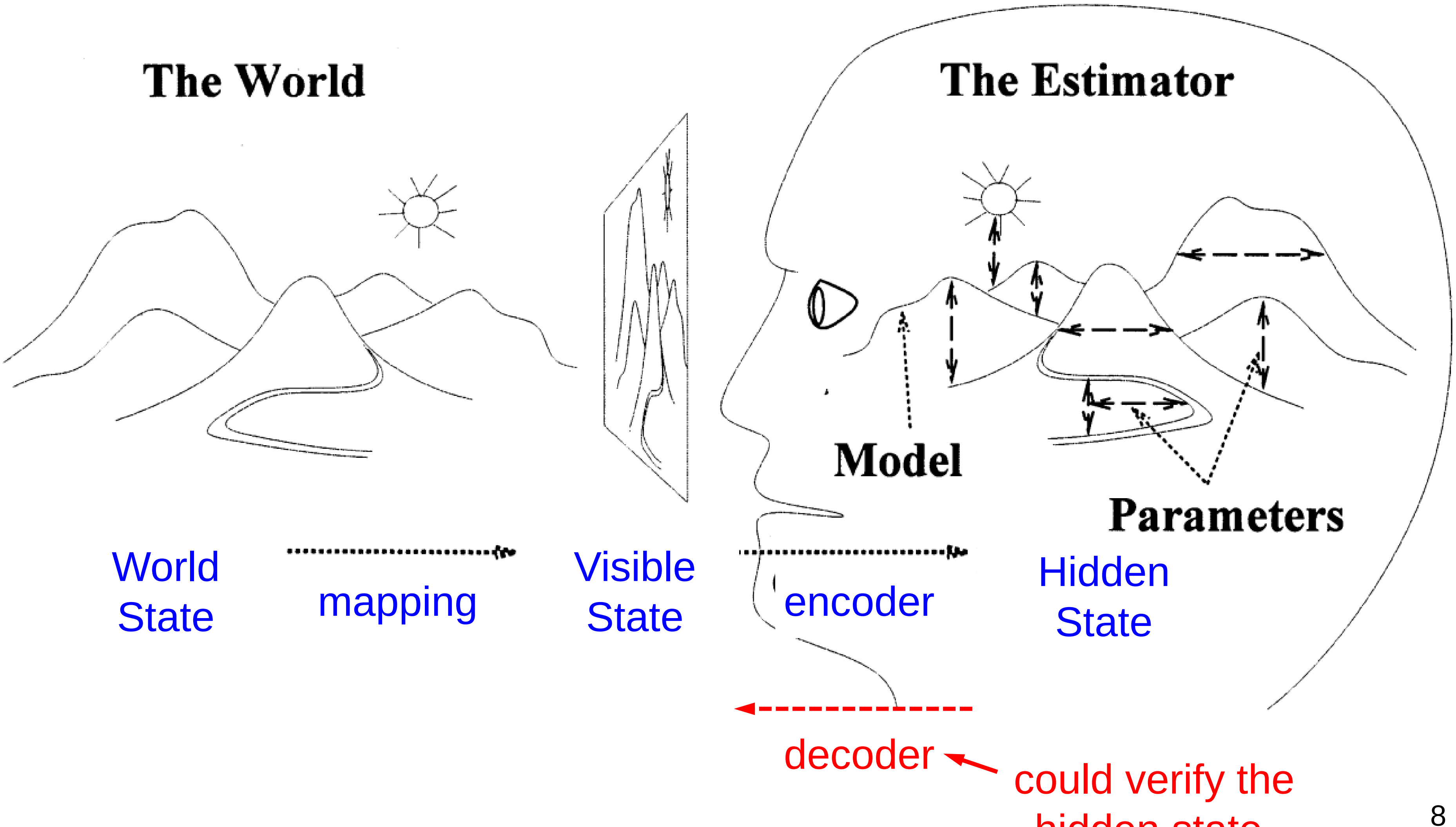


- Mostly supervised learning
  - Limits: availability of labelled data
- Remedy: unsupervised learning
  - particularly for lower layers
  - guided by findings from biology
- Unsupervised – how?

Overview

- **Hierarchical visual system**
- **Generative auto-encoder architectures**
  - MLP backpropagation
  - Transposed weights model
  - Helmholtz machine
- **Constraints**
  - few hidden neurons
  - weight constraints
  - sparse hidden activations
  - non-negativity
  - denoising
  - other

Generative Model

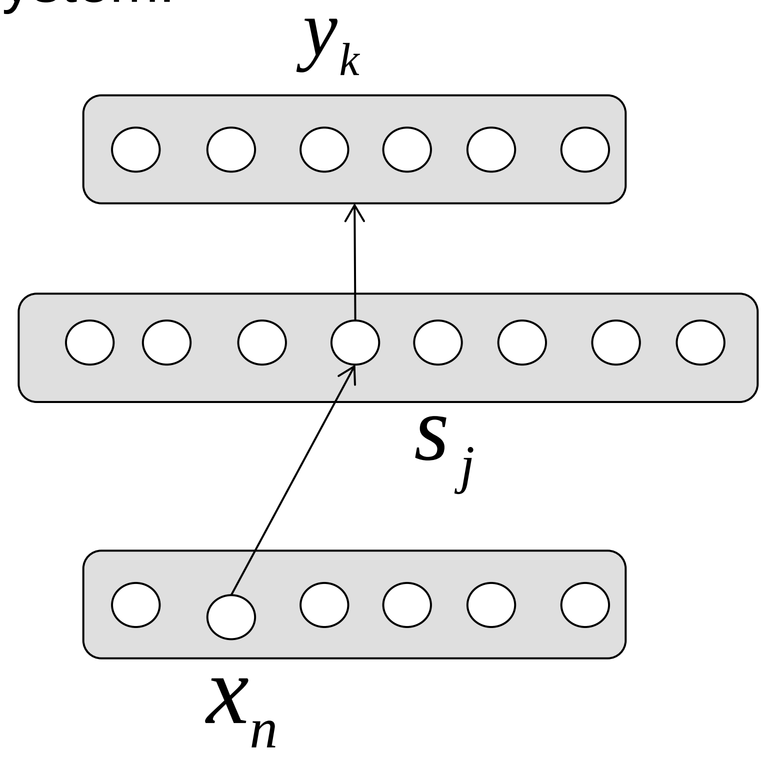


Overview

- **Hierarchical visual system**
- **Generative auto-encoder architectures**
  - MLP backpropagation
  - Transposed weights model
  - Helmholtz machine
- **Constraints**
  - few hidden neurons
  - weight constraints
  - sparse hidden activations
  - non-negativity
  - denoising
  - other

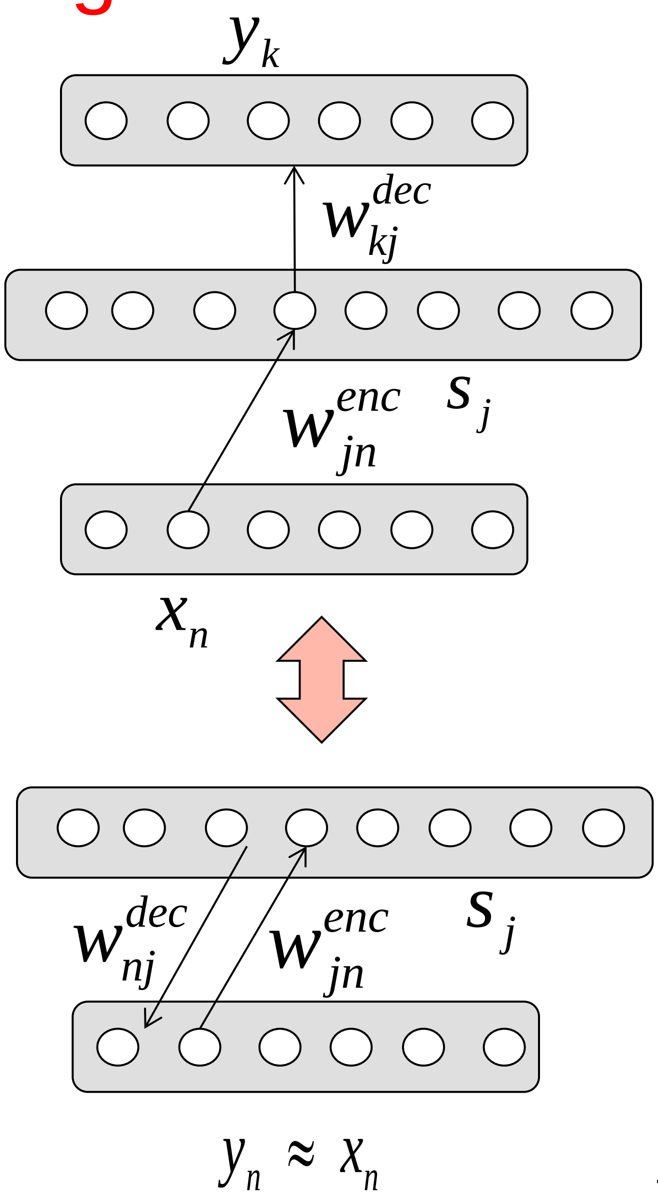
MLP / Error Backpropagation

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

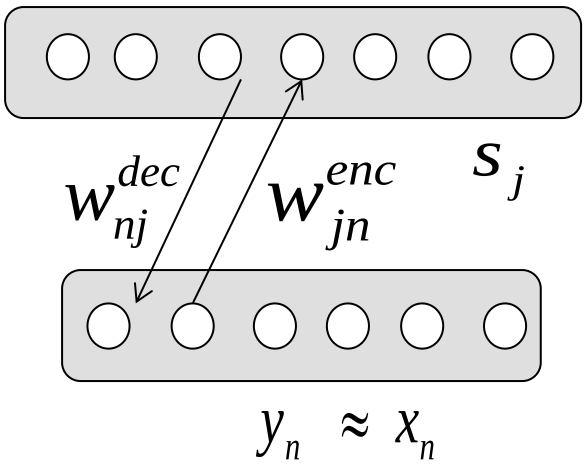


MLP / Error Backpropagation

- Special case of MLP: auto-encoder
- Training
  - Task: $y \overset{!}{\approx} x$
  - "Unsupervised" since no extra labels
  - Often just one hidden layer
  - Or, may be a deep architecture
- After training
  - The hidden code is of interest!



Error Minimization

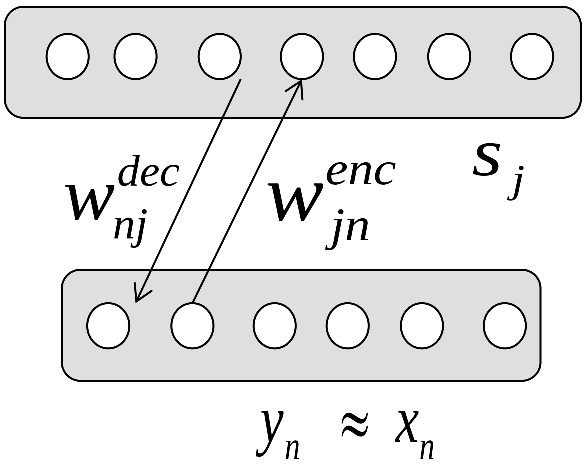


- "Feedforward" activation:
  - Encoder: $\vec{s} = \underset{\text{transfer function}}{f}(W^{enc}\vec{x})$
  - Decoder: $\vec{y} = W^{dec}\vec{s}$
- Error function: $E(W,\vec{s}) = \sum^{data} \frac{1}{2}(\underbrace{\vec{x}-\vec{y}}_{\vec{e}})^2$
  - input data is the target
- Learning: $\Delta W^{dec} \approx -\frac{\partial E}{\partial W^{dec}} = \vec{e}\vec{s}$
  - slightly modify weights for each data point, using the error
  - Error-backpropagation to train W^{enc}

Overview

- **Hierarchical visual system**
- **Generative auto-encoder architectures**
  - MLP backpropagation
  - Transposed weights model
  - Helmholtz machine
- **Constraints**
  - few hidden neurons
  - weight constraints
  - sparse hidden activations
  - non-negativity
  - denoising
  - other

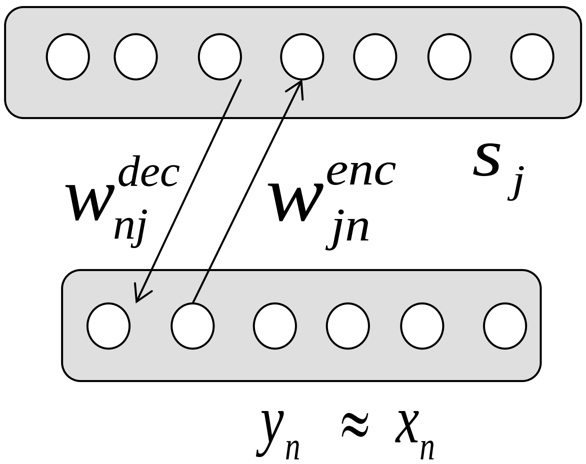
Transposed Weights Model



- Disadvantage of MLP:
  - Need error-backpropagation for W^{enc}
  - Training of W^{dec} was OK

- Solution:
  - Train only W^{dec}
  - Set: $W^{enc} = (W^{dec})^T$
  - No backpropagation required
  - Weights will scale to get the reconstruction right
  - Still, biologically unrealistic

Interpretation of Weights



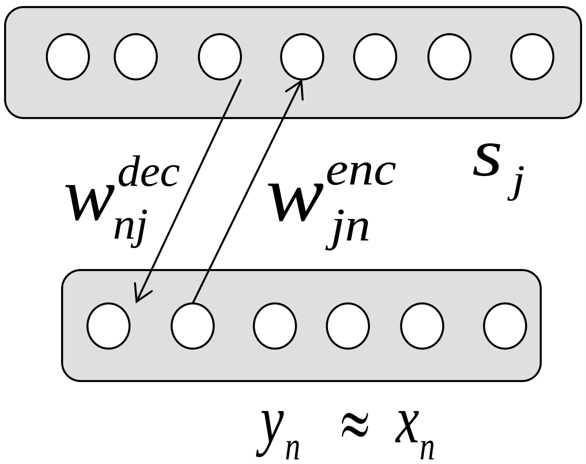
$$\vec{y} = W^{dec}\vec{s} = \sum_{j}^{N^{hidden}} \underbrace{\vec{w}_j^{dec}}_{\text{basis functions}} s_j$$

- The reconstructed vector is a superposition of the outgoing vectors (“basis functions“) of the hidden neurons
- The contributions of these functions is scaled by the hidden neuron activities
- If $W^{enc} = (W^{dec})^T$ then basis function ~ receptive field of a neuron
- A generative model decomposes its inputs into basis functions, which are often independent and which might represent meaningful components of the world

Overview

- **Hierarchical visual system**
- **Generative auto-encoder architectures**
  - MLP backpropagation
  - Transposed weights model
  - Helmholtz machine
- **Constraints**
  - few hidden neurons
  - weight constraints
  - sparse hidden activations
  - non-negativity
  - denoising
  - other

Helmholtz Machine



- Alternative to backpropagation or to transposing the weights
- Algorithm: Wake-sleep algorithm.

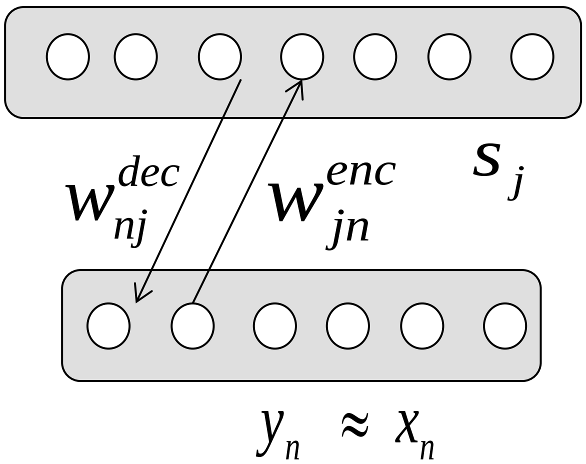
- *Wake phase* learning step for *W*^{dec} (as previously):

$$\Delta W^{dec} \approx e \cdot s$$

- The *sleep phase* for *W*^{enc} turns the model upside down:
  - Generate random activities $\tilde{s}$ *(training "data")*
  - From these, generate "imagined" inputs: $\tilde{x} = W^{dec}\tilde{s}$
  - Sleep phase learning step:

$$\Delta W^{enc} \approx (\tilde{s} - W^{enc}\tilde{x})\tilde{x}$$

Generative Models



- A perfect reconstruction of the input could be acieved trivially as

$$W^{enc} = W^{dec} = \mathrm{I} \quad \text{(identity matrix)}$$

and with linear units.

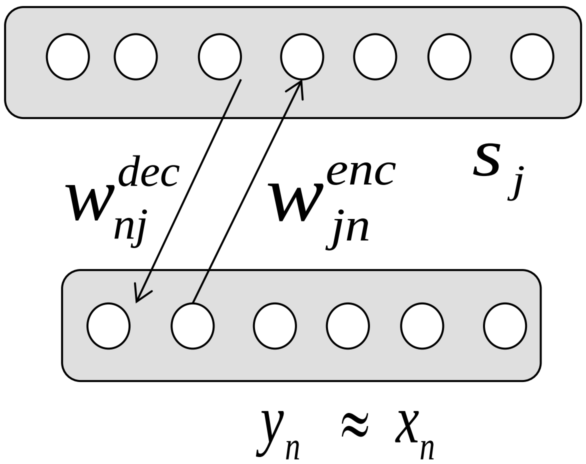
$\rightarrow$ No extraction of interesting features from the data!

- We should apply some contraints:
  - let the hidden layer re-code the data in interesting ways

Overview

- **Hierarchical visual system**
- **Generative auto-encoder architectures**
  - MLP backpropagation
  - Transposed weights model
  - Helmholtz machine
- **Constraints**
  - few hidden neurons
  - weight constraints
  - sparse hidden activations
  - non-negativity
  - denoising

Constraints on Generative Models



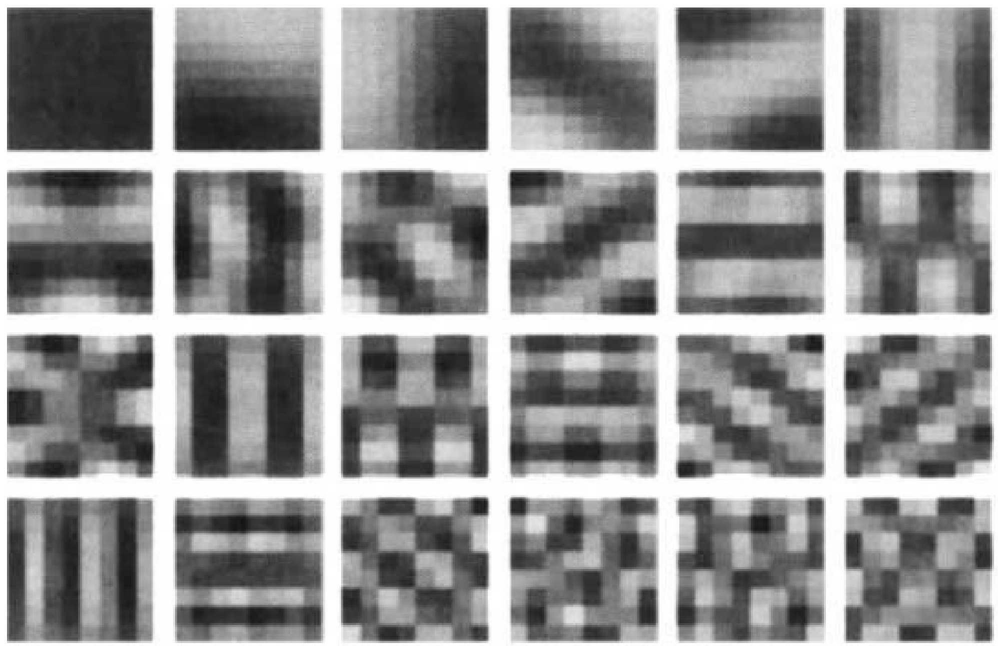
- Additional cost terms / “constraints” lead to interesting coding:
- On the *structure*:
  - few hidden neurons ~ PCA
  - weight constraints → model of retinal ganglion cells
- On the *code*:
  - sparse hidden activations → model of V1 edge detector cells
- On *structure and code*:
  - non-negative matrix factorization → part-based coding
- On the *data*:
  - denoising autoencoder

Overview

- **Hierarchical visual system**
- **Generative auto-encoder architectures**
  - MLP backpropagation
  - Transposed weights model
  - Helmholtz machine
- **Constraints**
  - few hidden neurons
  - weight constraints
  - sparse hidden activations
  - non-negativity
  - denoising
  - other

Neural Principle Component Analysis

- Bottleneck: *few* hidden linear neurons



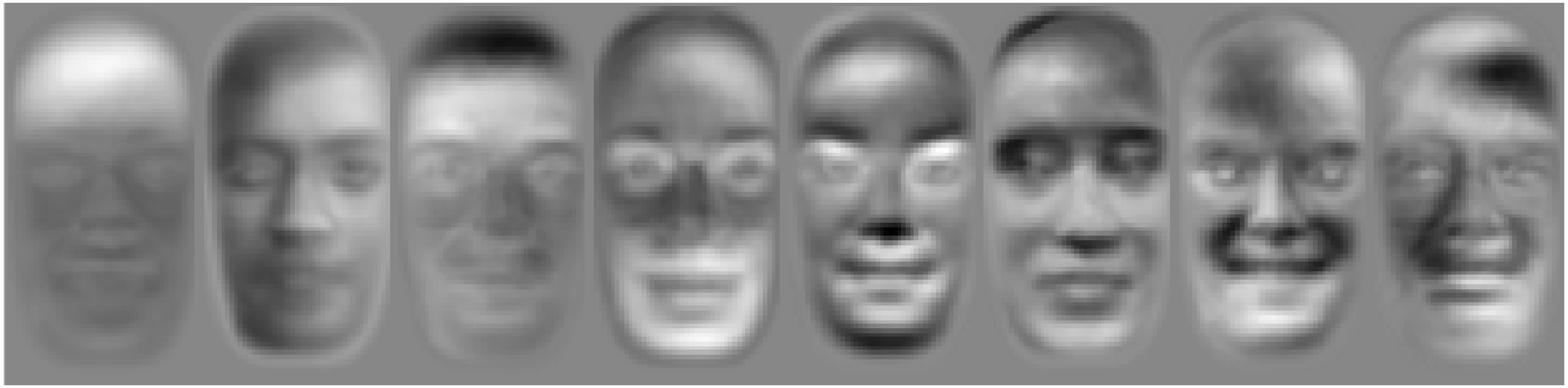
W^{enc}

- Training data: patches of grey-scale natural images
- Resulting basis functions span subspace with large variance
- Same subspace discovered as by PCA
- Sanger's rule finds 1^{st} PC first, then 2^{nd} PC, and so on

Neural PCA

- Bottleneck: *few* hidden linear neurons

W^{enc}



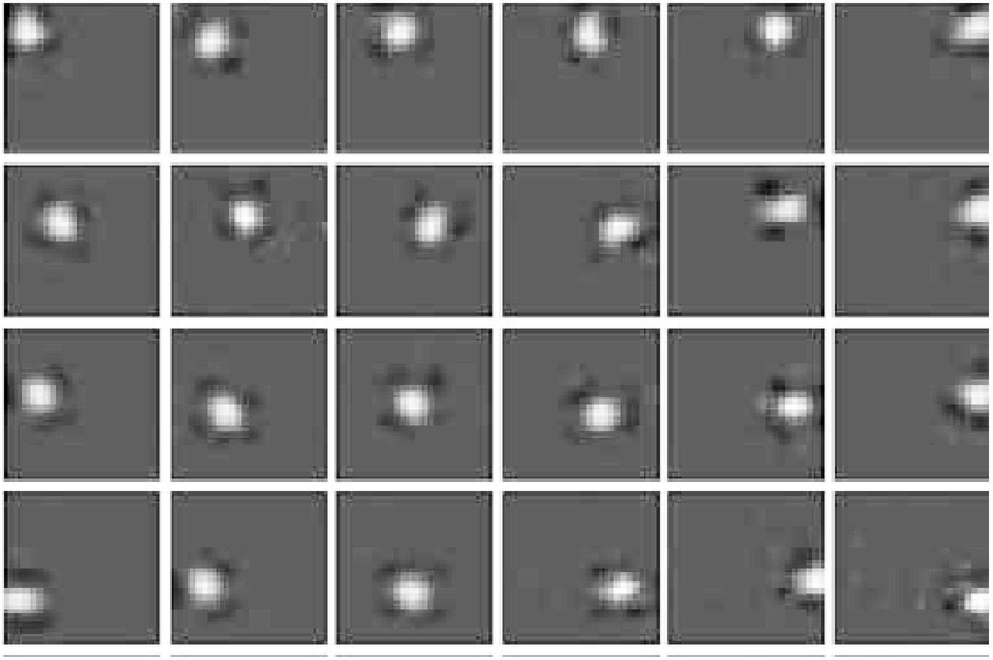
- Training data: grey scale images of faces, centred
- Resulting components show to direction of largest variance - *Eigenfaces*

Overview

- **Hierarchical visual system**
- **Generative auto-encoder architectures**
  - MLP backpropagation
  - Transposed weights model
  - Helmholtz machine
- **Constraints**
  - few hidden neurons
  - weight constraints
  - sparse hidden activations
  - non-negativity
  - Denoising
  - other

Weight Constraint

- Imposing a soft weight constraint (indirectly reduces firing rate)



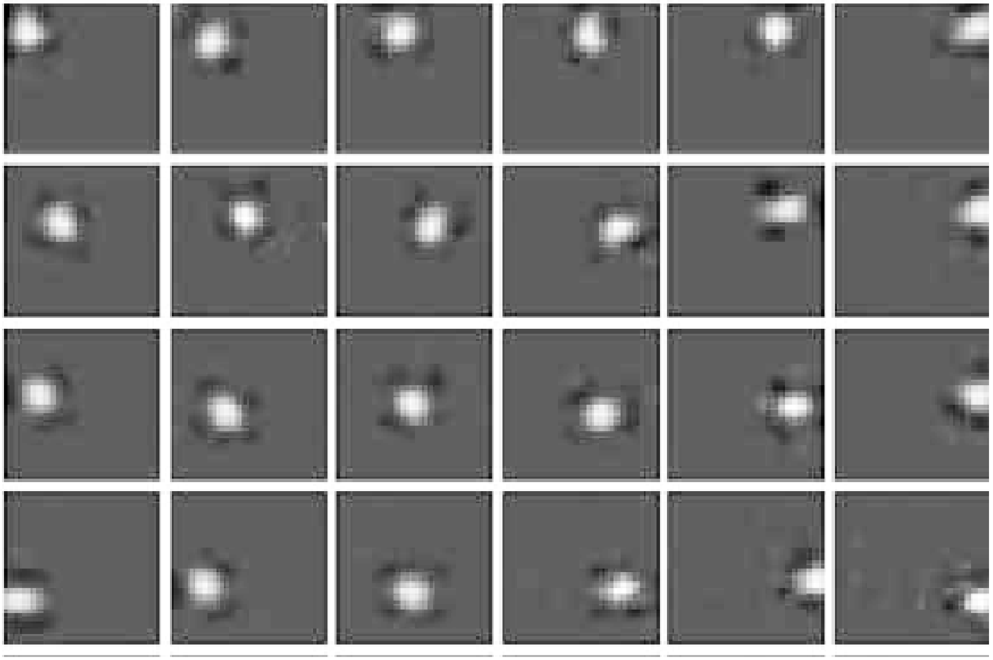
W^{enc}

- Training data: randomly chosen natural image patches
- Resulting receptive fields have centre-surround structure like retinal ganglion cells

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

Weight Constraint

- Implementation of the weight constraint (units are linear)



W^{enc}

L1 regularizer

$$\Delta w_{ij} \approx \underbrace{e_j \cdot s_i}_{\textit{good}\ \text{reconstruction}} \underbrace{- const_1 \cdot sign(w_{ij})}_{\text{regularization term}}$$

L2 regularizer

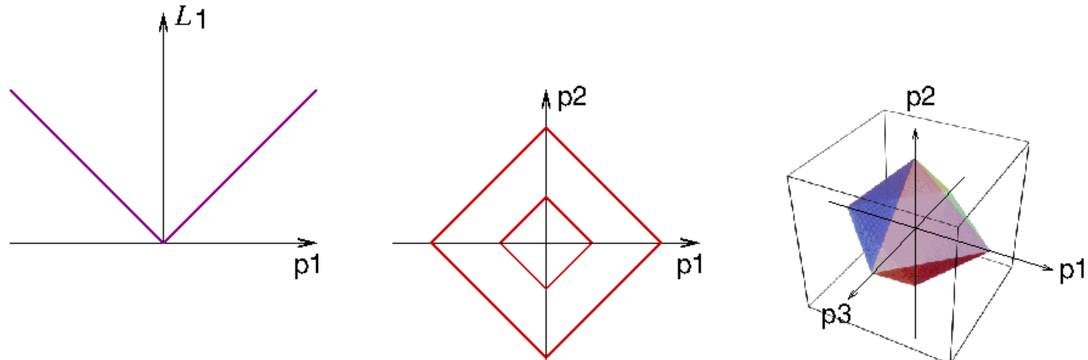
Here: $-\frac{\partial}{\partial w}|w| = -sign(w)$ as opposed to $-\frac{\partial}{\partial w} w^2 = -w$

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

Constraints with L1 or L2 Norm

- L1 Norm

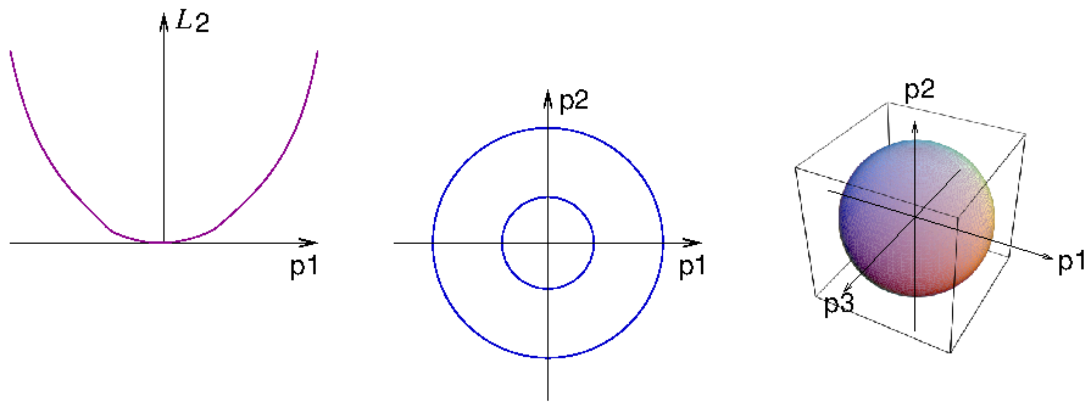
$$\| \vec{p} \|_1 = \sum_i | p_i |$$



$\rightarrow$ L1 norm favours sparse parameters $\vec{p}$ (weights)

- L2 Norm

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

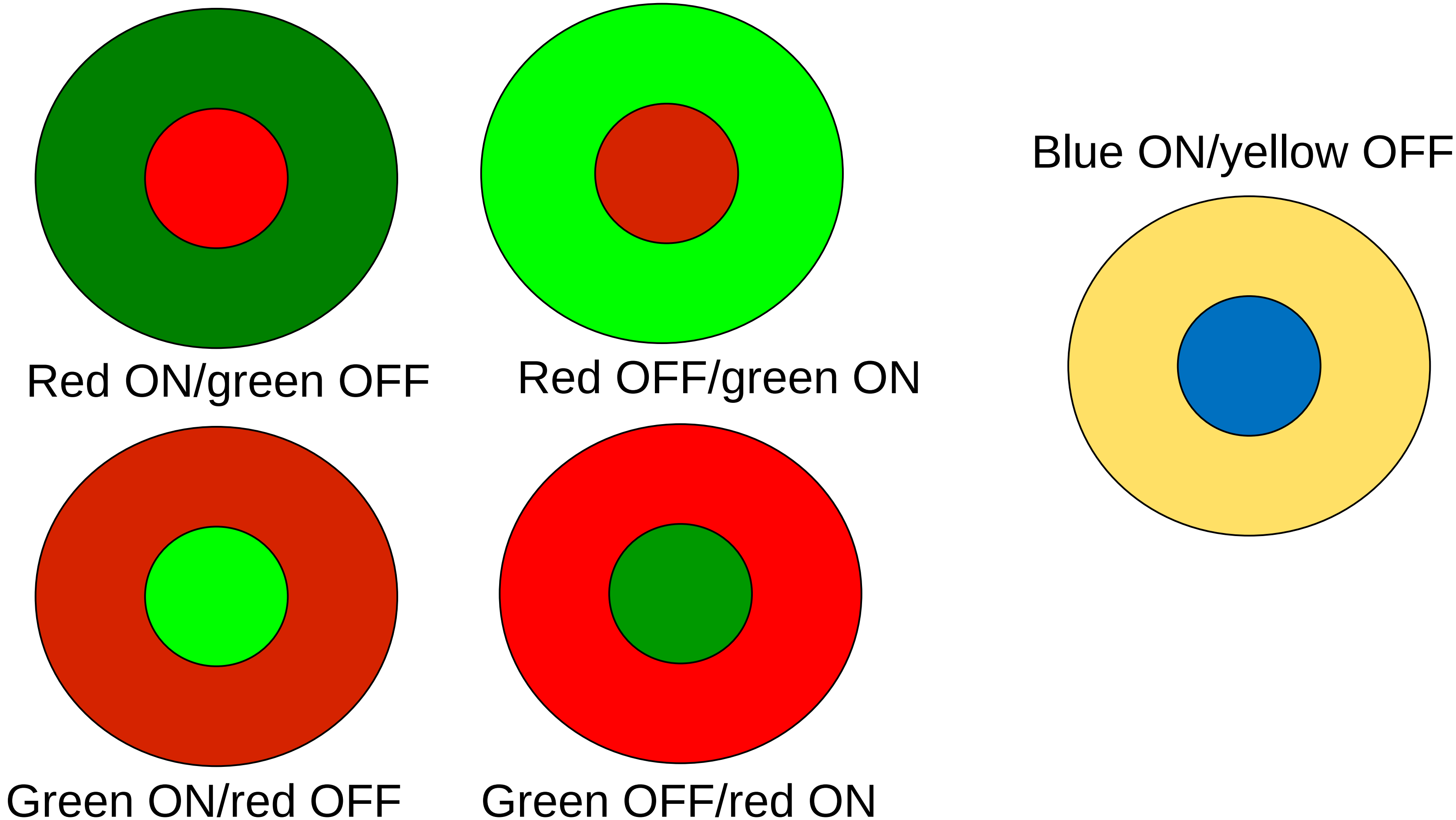


$\rightarrow$ L2 norm penalises large parameters, but will not rotate $\vec{p}$

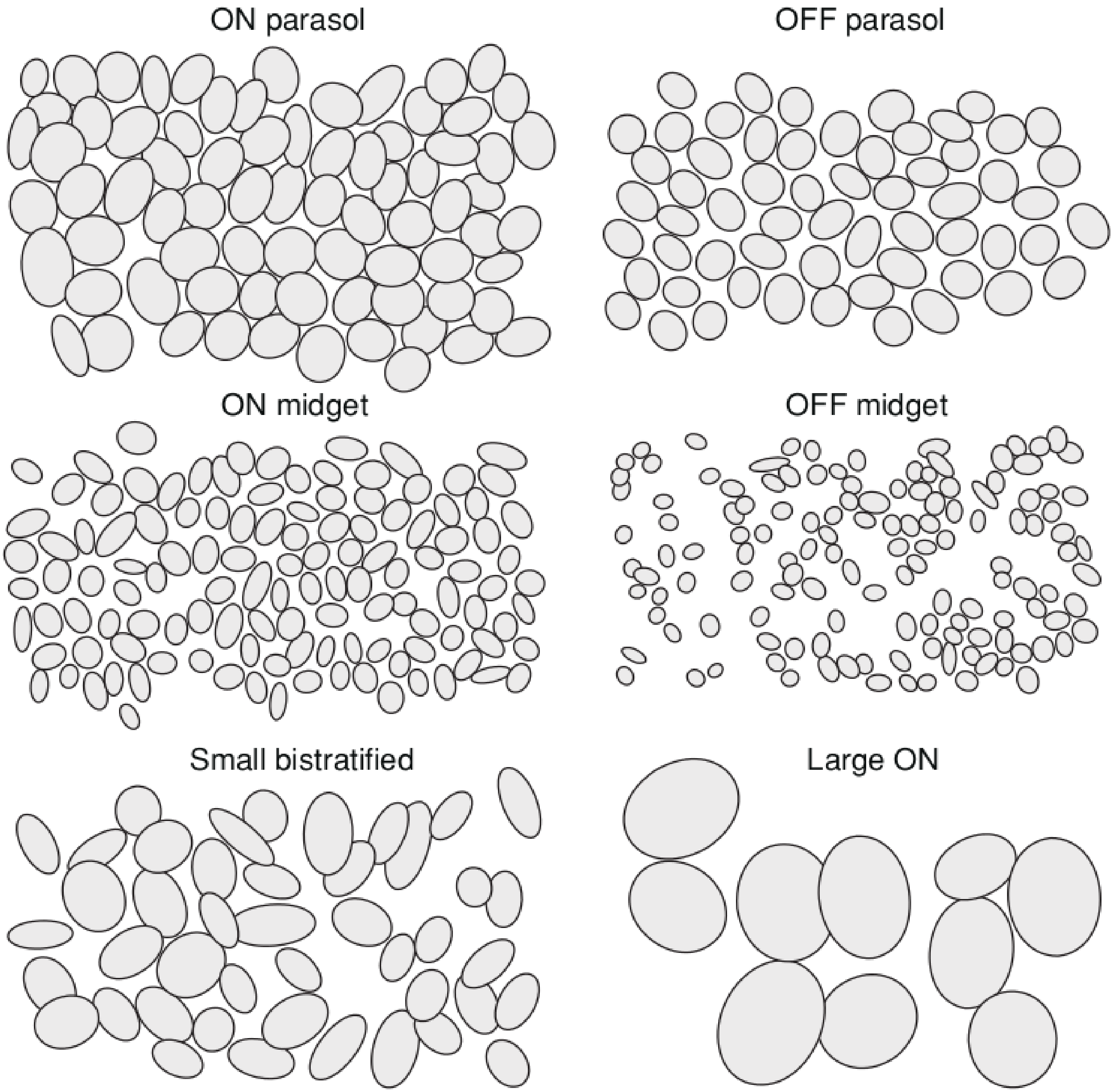
$\| \vec{p} \|_\infty = \max_i \{| p_i |\} \rightarrow L\infty$ norm favours all p_i to be the same

Weight Constraint

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



Weight Constraint



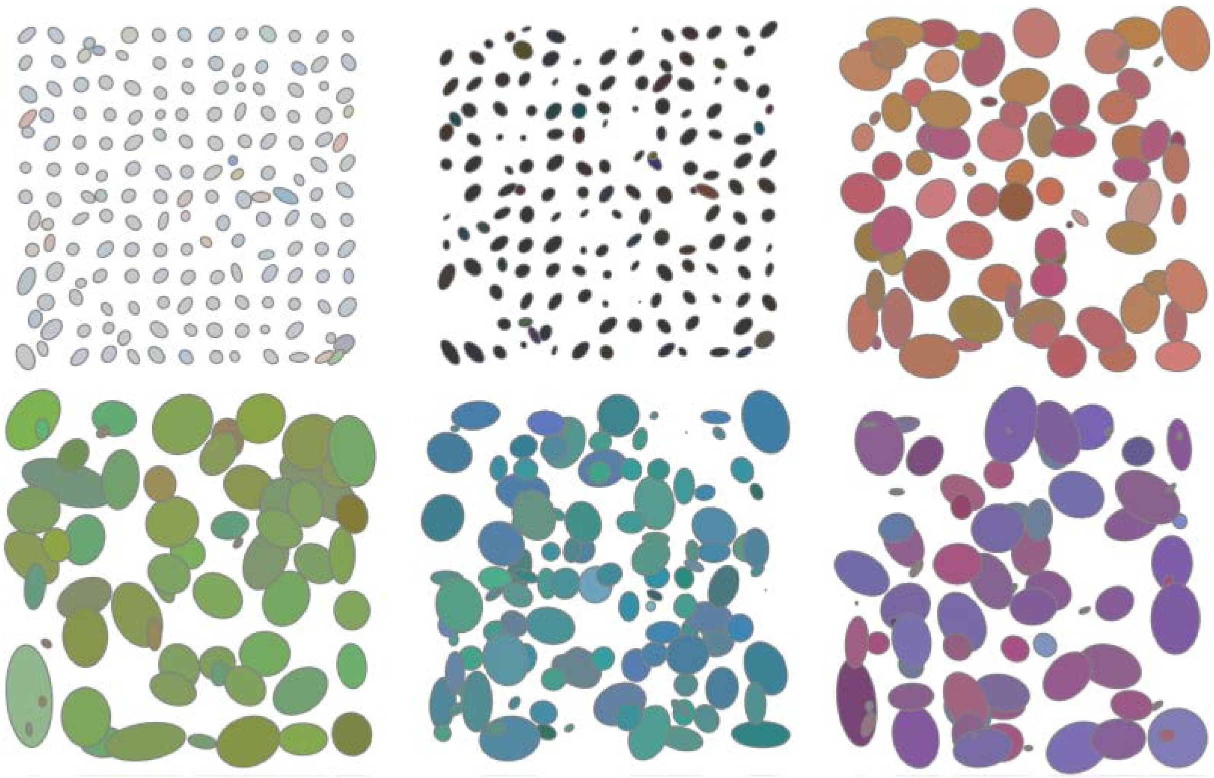
luminosity, phasic

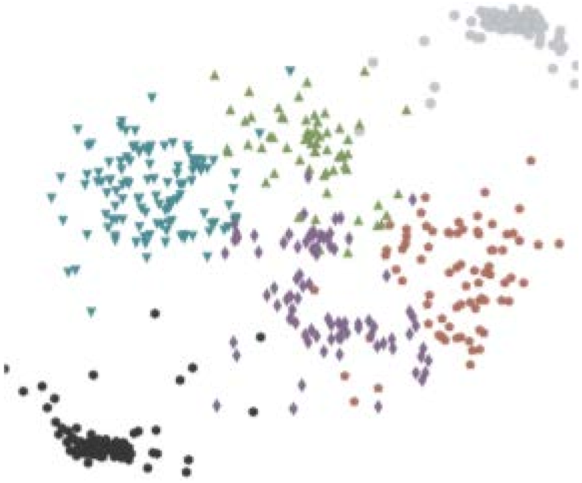
red-green

left: blue-yellow

Weight Constraint

- Training data: randomly chosen natural colour image patches

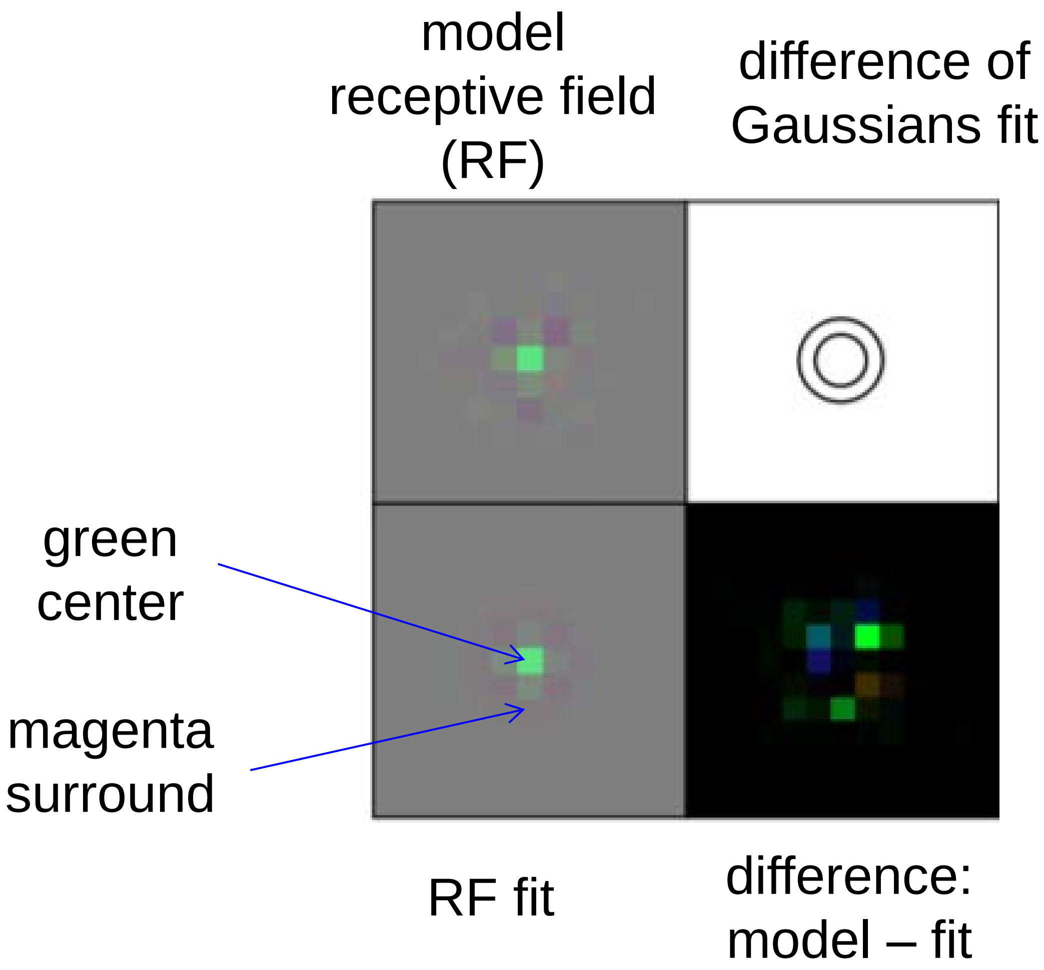




6 clusters

Weight Constraint

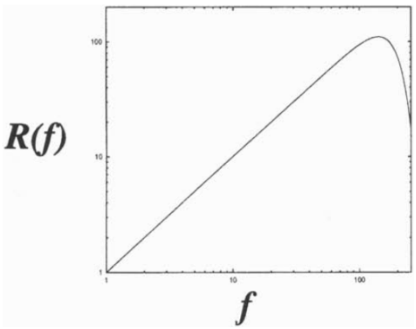
- Example cell from the “green” cluster



Overview

- **Hierarchical visual system**
- **Generative auto-encoder architectures**
  - MLP backpropagation
  - Transposed weights model
  - Helmholtz machine
- **Constraints**
  - few hidden neurons
  - weight constraints
  - sparse hidden activations
  - non-negativity
  - Denoising
  - other

Retinal Preprocessing



- Pre-processing of input images by filtering
- Filter in spatial frequency space

$$R(f) = f \cdot e^{-(f/const)^4}$$

has two terms:

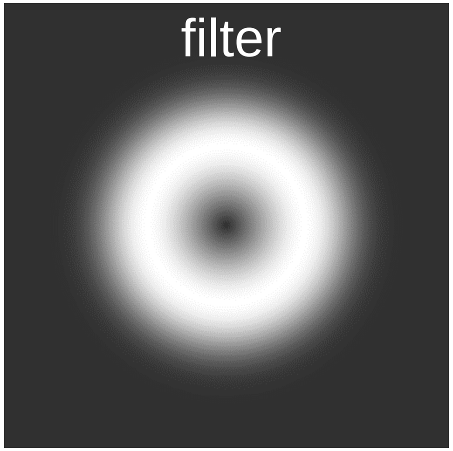
- f term reduces *low* frequencies

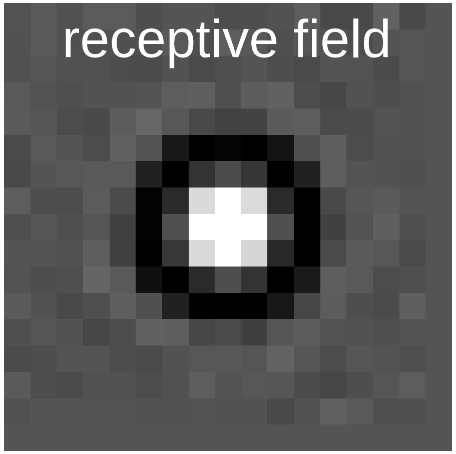
  $\rightarrow$ equalizes amplitude spectrum of $1/f$

- $e^{-(f/const)^4}$ term reduces *high* frequencies

  $\rightarrow$ reduces pixel noise

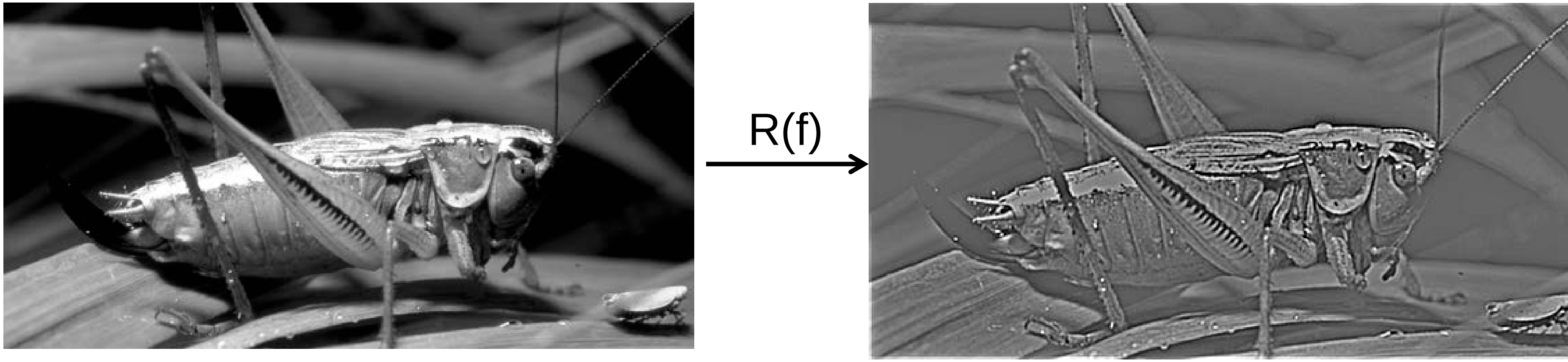
- Filter in image space resembles retinal ganglion cell receptive fields



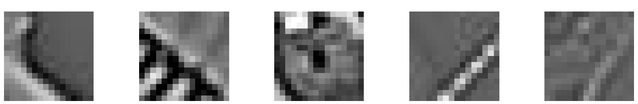


Olshausen, Field. Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1? Vis. Res., 1997

Retinal Preprocessing



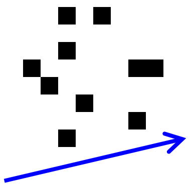
- Training data: random image patches (to match the input layer size) from filtered images:



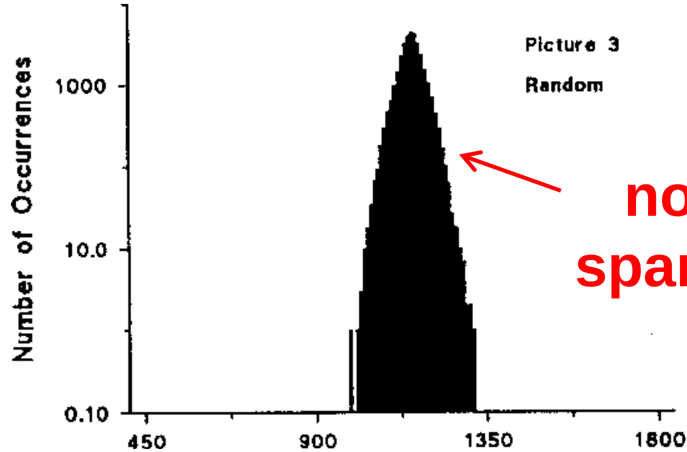
…

Sparse Coding

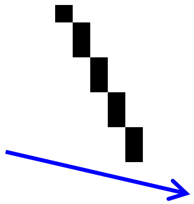
Centre-surround filtered image



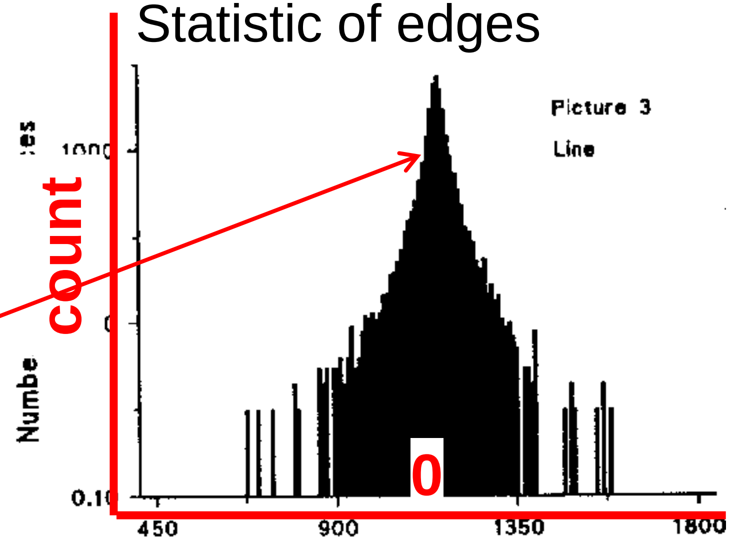
Statistic of random patterns



not sparse



Statistic of edges



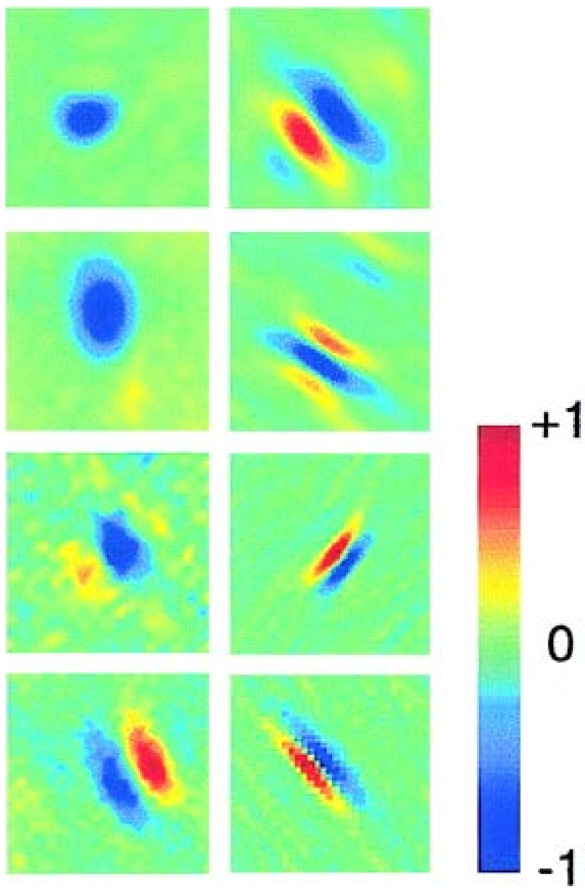
sparse

neuron activity

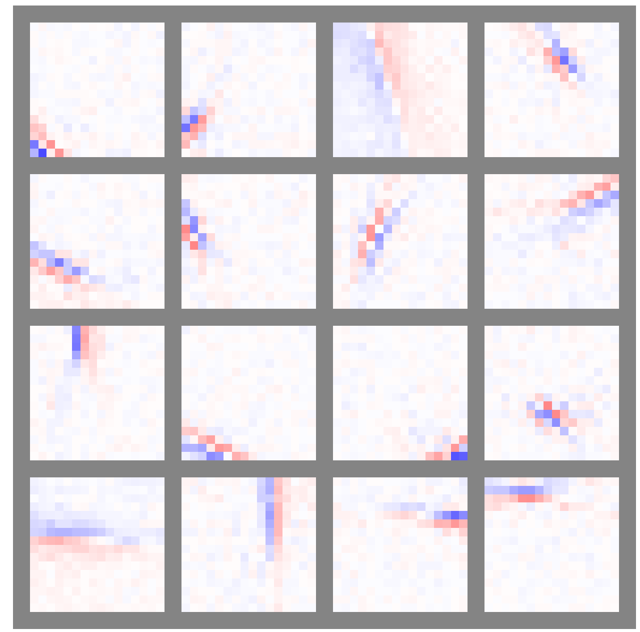
Barlow. Large Scale Neuronal Theories of the Brain., ch. What is the computational Goal of the Neocortex?, 1994

Sparse Coding

Primary visual cortex (V1) cell receptive fields (RF)



Selected trained RFs (from an overcomplete set)



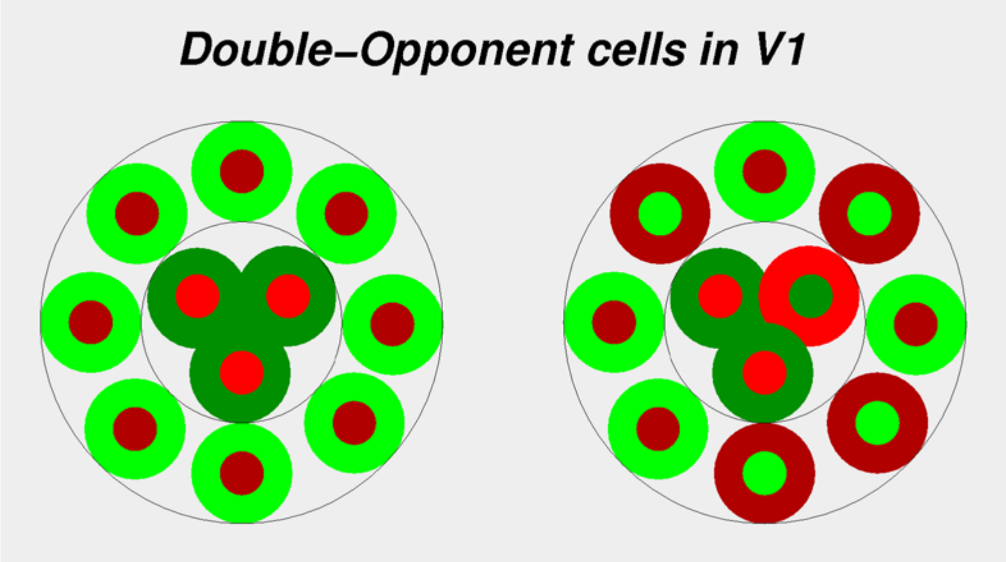
W^{enc}

- Constraint: sparse coding
- Resulting RFs are localized edges

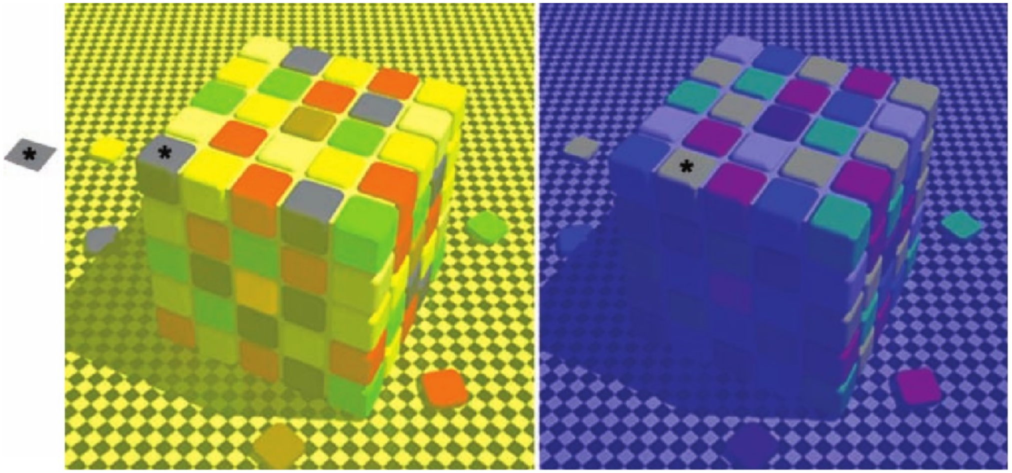
Ringach. Spatial Structure ... of Simple-Cell Receptive Fields in Macaque Primary Visual Cortex. J Neurophysiol. 2002

Sparse Coding

possible results for
color images



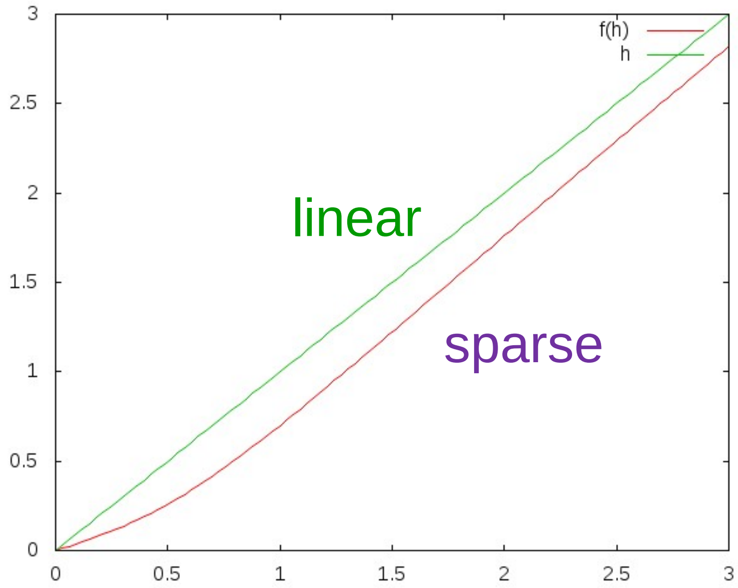
?
⇒ color constancy



Sparse Coding

- Sparse transfer function on the hidden layer, e.g.

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



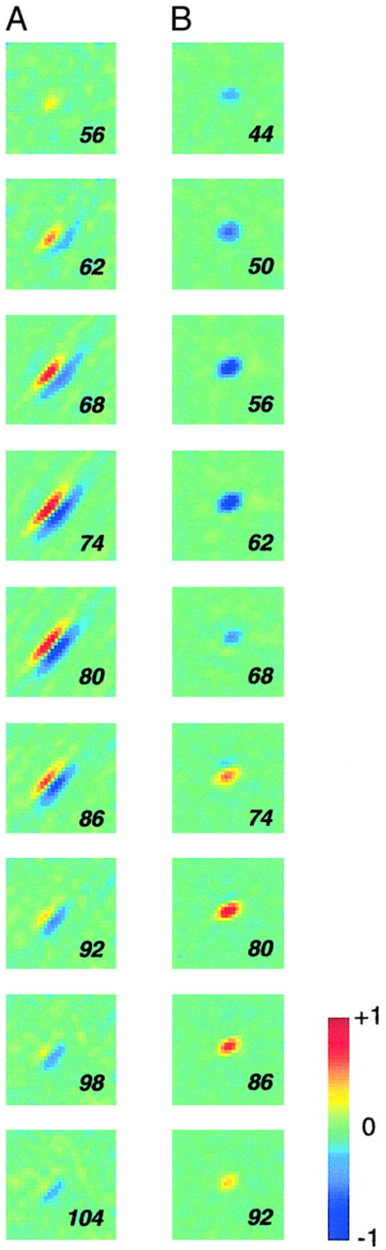
  - Reduces small activations but retains large activations
  - A weight decay (regularization) term will be needed to keep weights small

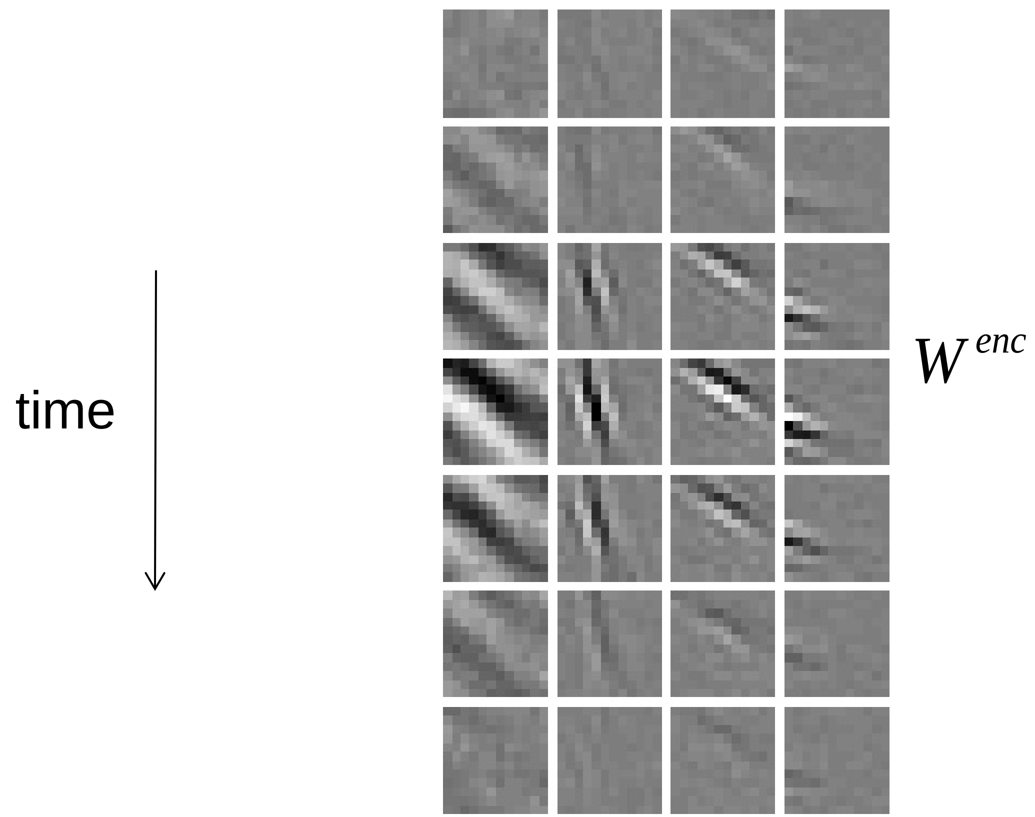
    (large weights would counteract sparseness)

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

Sparse Coding

generative sparse model applied to movies
→ spatiotemporal V1 receptive fields



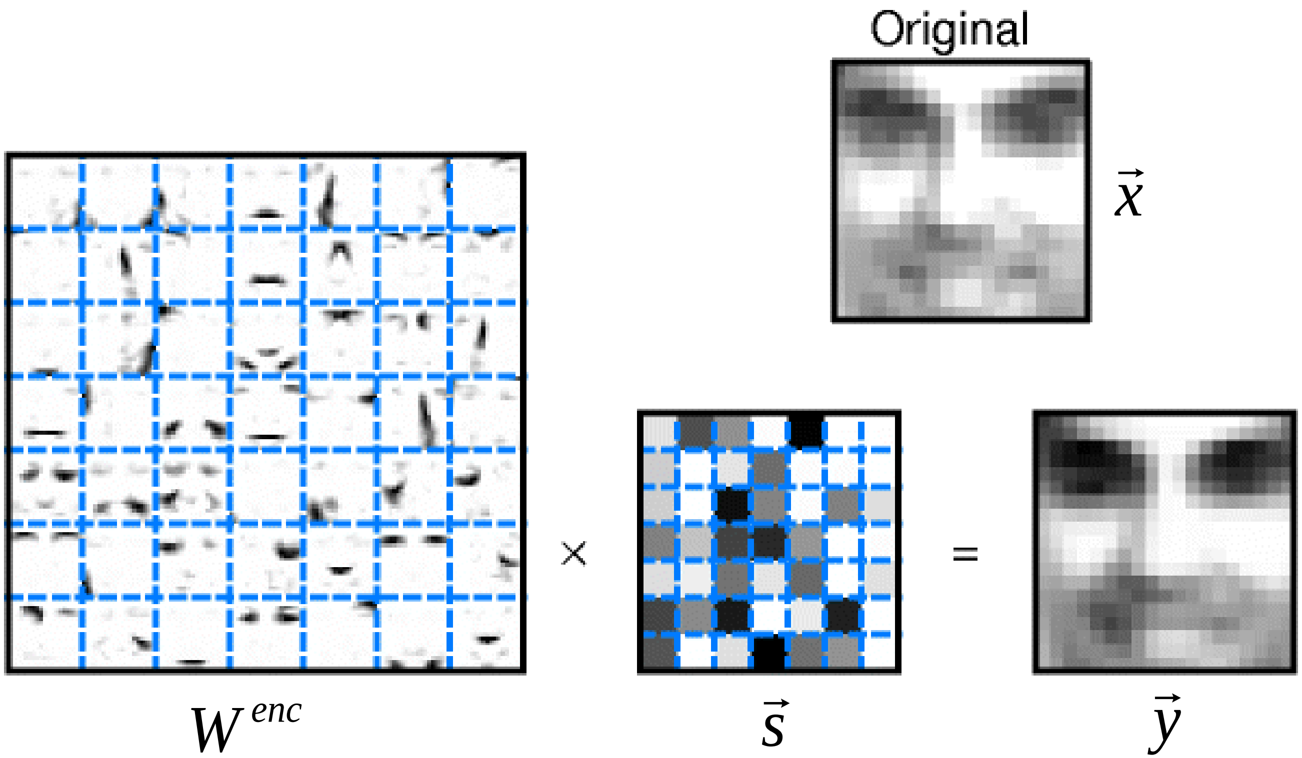


Overview

- **Hierarchical visual system**
- **Generative auto-encoder architectures**
  - MLP backpropagation
  - Transposed weights model
  - Helmholtz machine
- **Constraints**
  - few hidden neurons
  - weight constraints
  - sparse hidden activations
  - non-negativity
  - denoising
  - other

Non-negative Matrix Factorization (NMF)

Constraint:
non-negativity of
all activations
and weights



- Training data: centered faces
  (white pixels encoded by zero activity; dark pixels by positive activations)
- Resulting basis functions: part-based representations

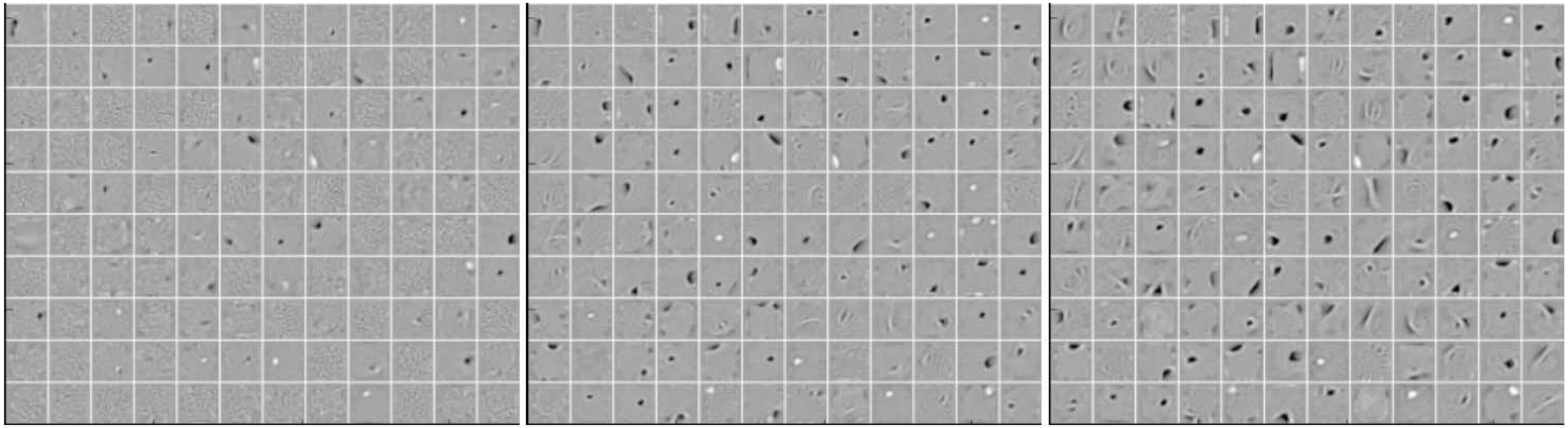
Lee, Seung. Learning the parts of objects by non-negative matrix factorization. Nature, 1999

Overview

- **Hierarchical visual system**
- **Generative auto-encoder architectures**
  - MLP backpropagation
  - Transposed weights model
  - Helmholtz machine
- **Constraints**
  - few hidden neurons
  - weight constraints
  - sparse hidden activations
  - non-negativity
  - denoising
  - other

Denoising Autoencoder

- Reconstruction from partially corrupted input patterns



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

W^{enc}

Training inputs: MNIST digits corrupted with zero-value “blank” pixels

Resulting basis functions: patchy, partially localised receptive fields

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

Overview

- **Hierarchical visual system**
- **Generative auto-encoder architectures**
  - MLP backpropagation
  - Transposed weights model
  - Helmholtz machine
- **Constraints**
  - few hidden neurons
  - weight constraints
  - sparse hidden activations
  - non-negativity
  - denoising
  - other

Newborn Looking Preferences to Faces

Table 1. *Number of babies who preferred each stimulus*

	Feature inversion		
Age	*Config*	Inversion	Neither
Newborns	9*	1	2
6-week-olds	0	0	12
12-week-olds	0	1	11

Model neuron RF

How explain neurons' innate complex shape preferences?

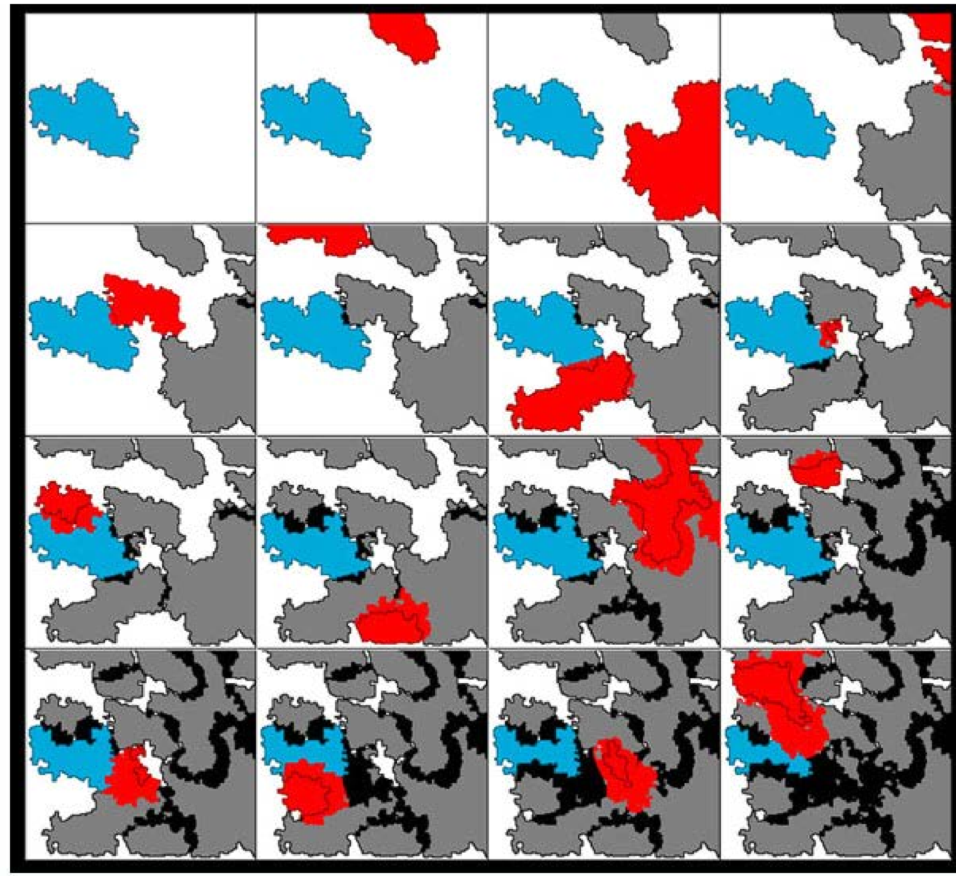
Model: newborn's face selective neurons have RFs in which weight patches are arranged in a *top-heavy triangle*.

Triangular prenatal activity patterns exist (for training)?

Data: Mondloch et al. Face Perception During Early Infancy. Psychol. Science, 1999
Model: Bednar, Miikkulainen. Learning Innate Face Preferences. Neural Computation, 2003

Data-Driven Prenatal Training

biological training data:
pre-natal retinal waves



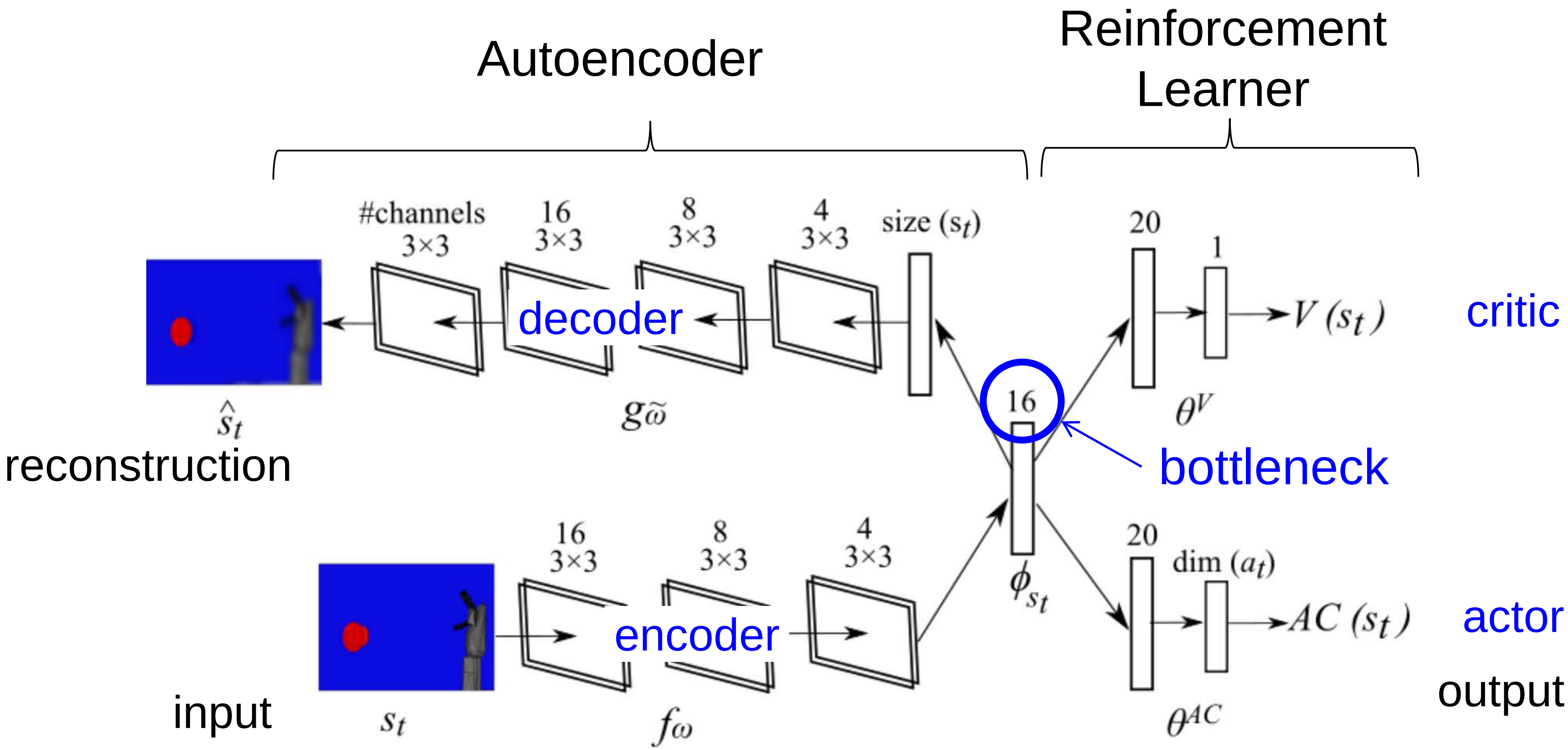
→ image-like properties:

- topographical relations
- edges
- convex figures
- coherent motion

note: "In the mature retina, the dendrites of On and Off ganglion cells are segregated into separate sublaminas of the inner plexiform la[illegible]nt these processes are multistratified ... ganglion cells with multistratified dendrites respond to the onset, as well as the offset, of light."

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

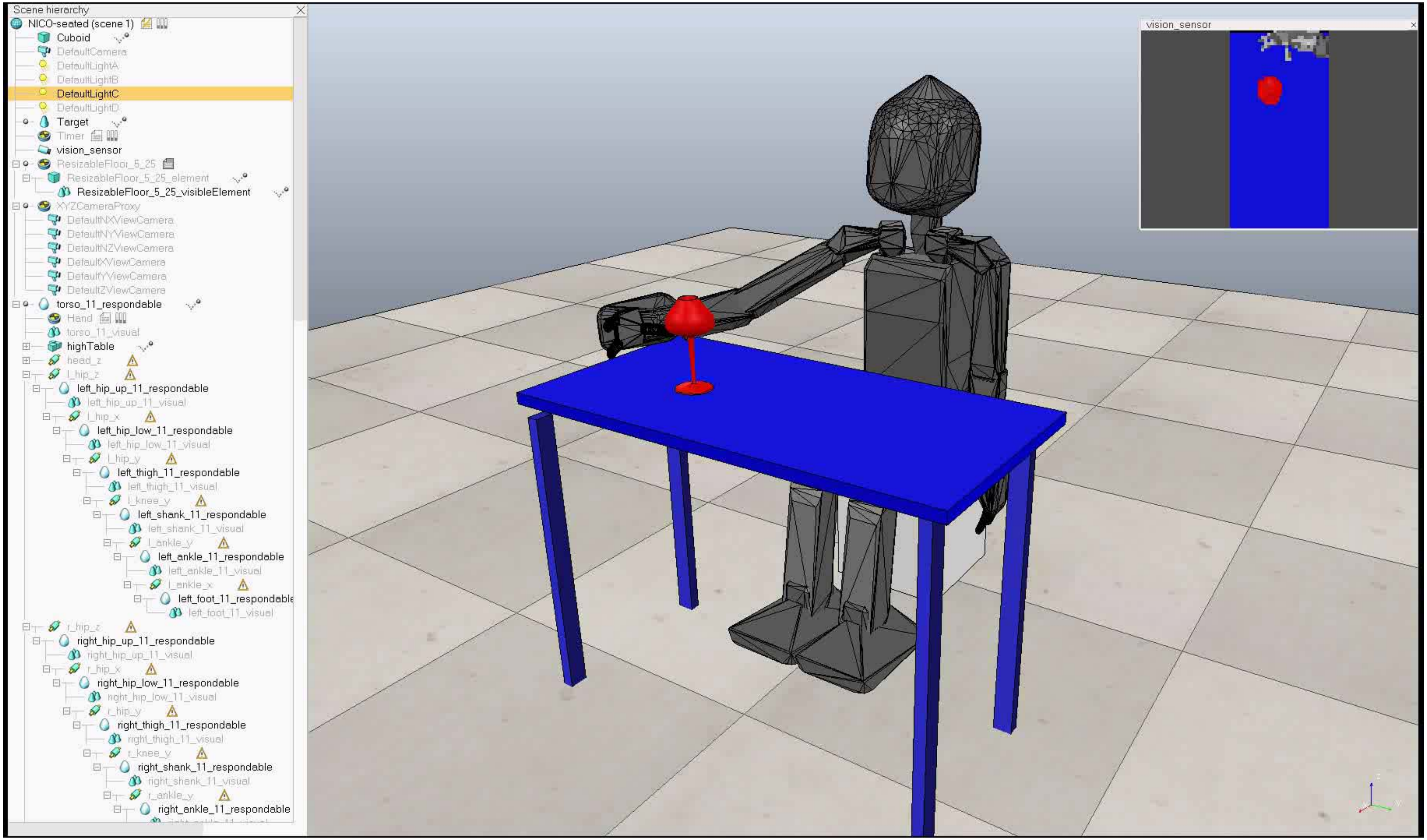
Example: Autoencoder as Preprocessor for a Reinforcement Learner



- Autoencoder learns many parameters of a deep visual NN → small reinforcement learner; partially shared parameters

Hafez, Weber, Kerzel, Wermter. Deep Intrinsically Motivated Continuous Actor-Critic for Efficient Robotic Visuomotor Skill Learning. In Preparation.

Example: Autoencoder as Preprocessor for a Reinforcement Learner



Hafez, Weber, Kerzel, Wermter. Deep Intrinsically Motivated Continuous Actor-Critic for Efficient Robotic Visuomotor Skill Learning. In Preparation.

Generative Models in Vision – Summary

- Hierarchical visual system with growing abstraction
- Weight matrices transform the representations
- Generative autoencoder models for unsupervised learning
- Various constraints on hidden encoding during learning:
  - few hidden neurons
  - weight constraints
  - sparse hidden activations
  - non-negativity
  - denoising
- More constraints needed to explain variety of cortical areas?
  - innate face preferences
  - peripheral / foveal preferences
  - slow / fast responses, …

Outlook

Next Lecture:

- 14th June: L09: Training Neural Networks

Seminar:

- 7th June 6pm: Draft paper deadline (today!)

Block seminar dates:

- Thu/Fri 19/20 July
- Mon/Tue 23/24 July

Oral exam dates (tentative): 8./9./10. Aug. & 26./27. Sept.