Introduction to Neural networks
(Psy5038W)

The visual system: overview of a large-scale neural architecture

Daniel Kersten
Computational Vision Lab
Psychology Department, U. Minnesota

kersten.org

Goals

- Provide an overview of a major brain subsystem to help anchor concepts in neural network theory.
- Behavioral, functional requirements that determine the computations that networks must do.
- Discuss issues of neural representation.
- Connect various parts and functions of the visual system with neural network ideas we've studied

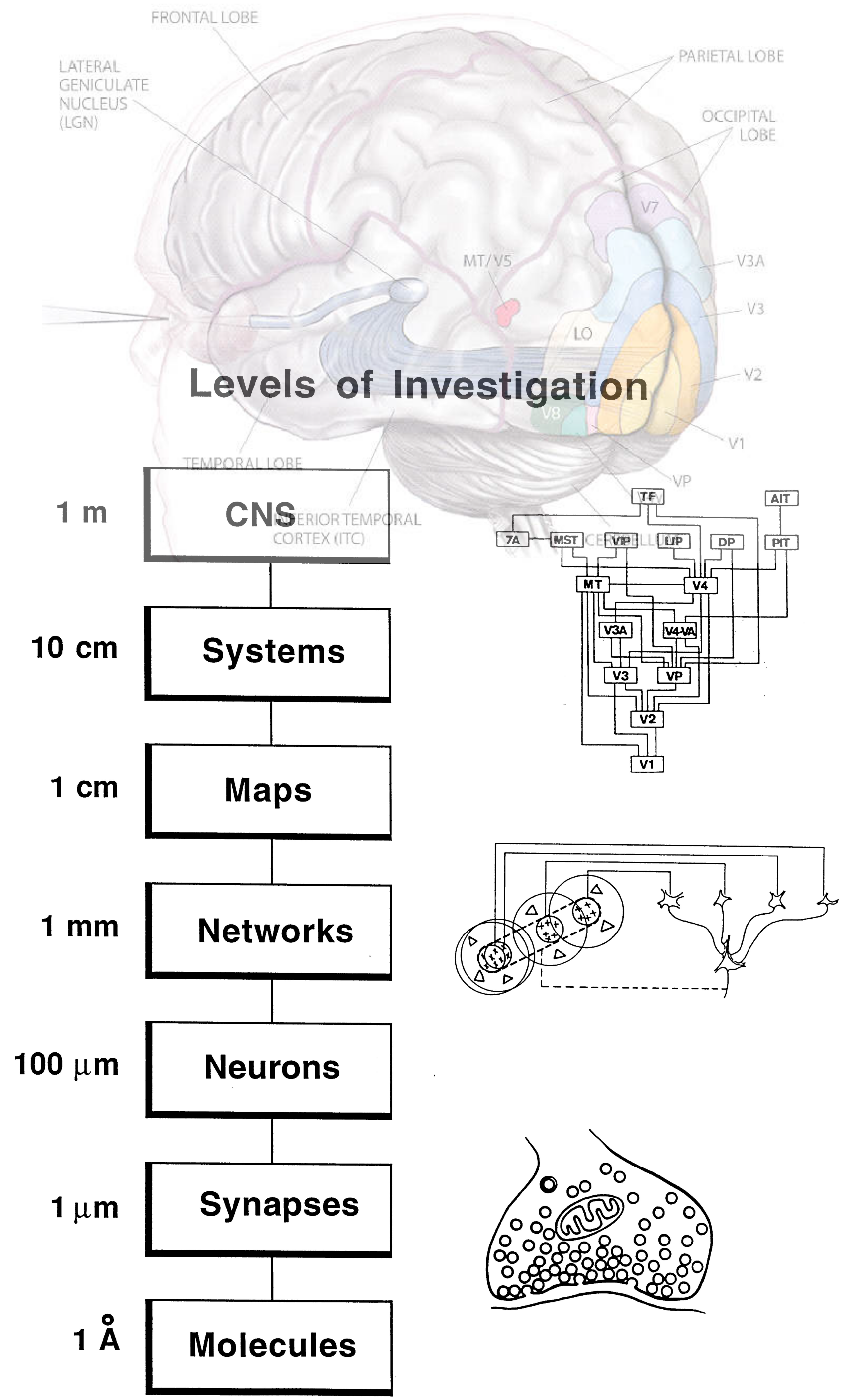
complex information processing system

where to start?

anatomy, neurophysiology… at what scale?

neurocomputational theory?

or behavior?



Visual behavior—jobs of vision

Within-object relations: Object perception

- feature grouping, categorization, identification
- object properties/attributes: size, shape,material, pose,expressions,…

Viewer-object relations

- navigation,heading ,time-to-contact,…
- manipulation/grasp
- tracking

Object-object relations

- relative depth, relative motion, scene interpretation, planning, scene recognition,…

Inferences range over:

- types of features & attributes (shapes, material)
- levels of abstraction (parts, objects, actions, scenes)
- spatial scales
- relationships

Descriptions are inferences of object properties and relationships — i.e. causes of image intensities, not of image intensity patterns

A working hypothesis is that these inferences are based on deep, generative knowledge of how virtually any natural image could be produced

computational problems

model uncertainty

vision is concerned with causes of image intensity patterns, but the causes of behavioral relevance are encrypted and confounded

many hypotheses about cause can be consistent with the same local image evidence

local variations in image evidence can be consistent with the same cause

accurate perceptual decisions resolve these ambiguities by combining lots of image evidence with built-in knowledge

rationale for Bayesian models of information processing

computational problems

scalability

Solving toy (low-dimensional) problems rarely scales up to deal with the complexity of natural images.

Humans have the capacity to deal with an enormous space of possible objects (30 to 300K) as they appear in different contexts in natural images for different tasks.

rationale for understanding feedforward architecture, e.g. "deep" convolutional networks

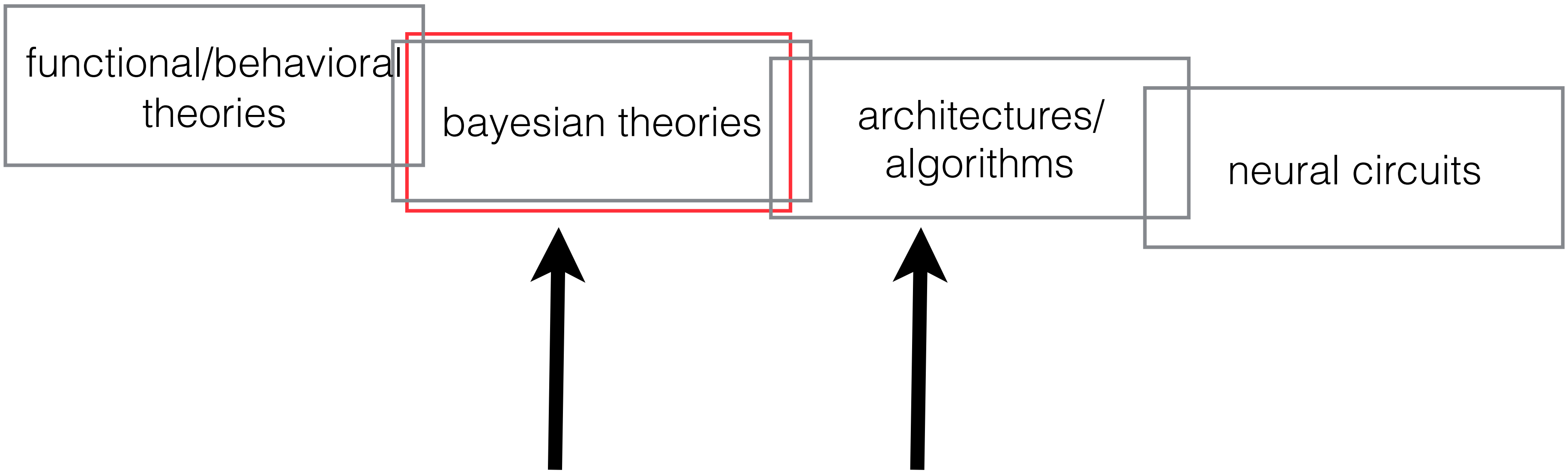
computational problems

task flexibility

Vision stimulates and support answers to a limitless range of questions. Human vision doesn't just recognize objects and patterns, it supports the interpretation of scenes.

rationale for a computational understanding of sequential processing and control…the role of feedback

levels of analysis

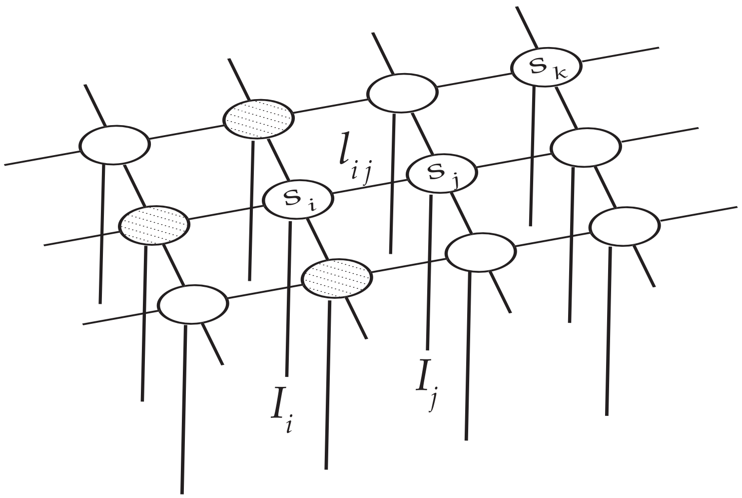


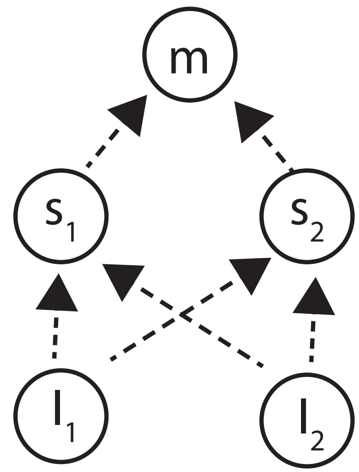
computational vision

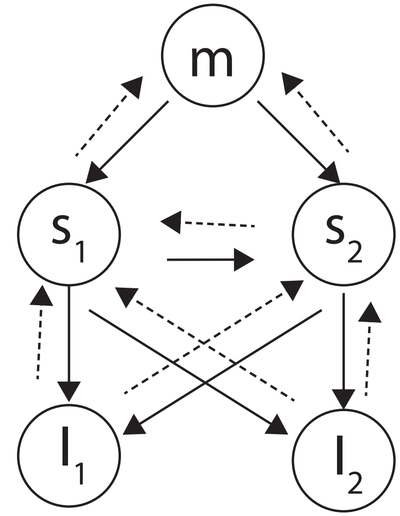
bayesian decision theory provides framework for modeling uncertainty

architectures/algorithms provide tools for understanding scalability and task flexibility

graphical models







lateral organization,

lateral processing, reciprocal interactions between features of similar type

hierarchical organization

feedforward processing, increasing abstraction

hierarchical organization: generative view

feedforward, feedback & lateral processing

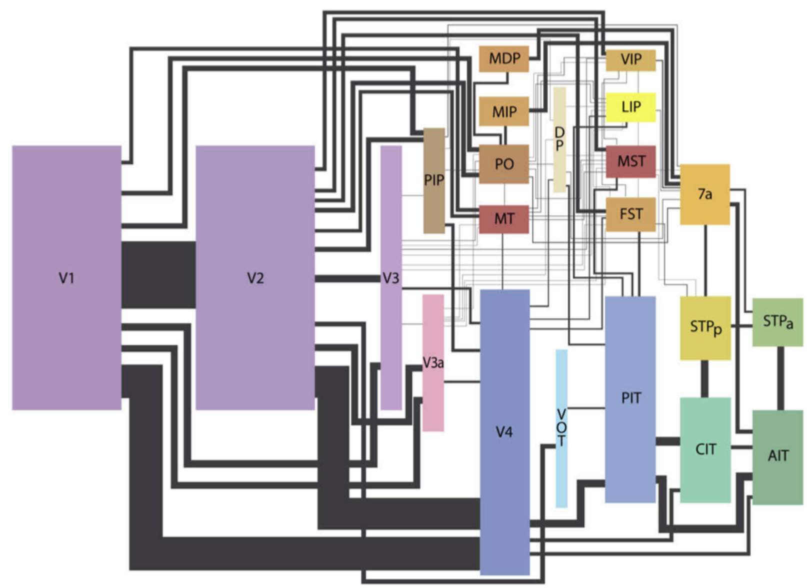
rationale for computations on undirected graphs

computations on directed graphs, e.g. mcmc

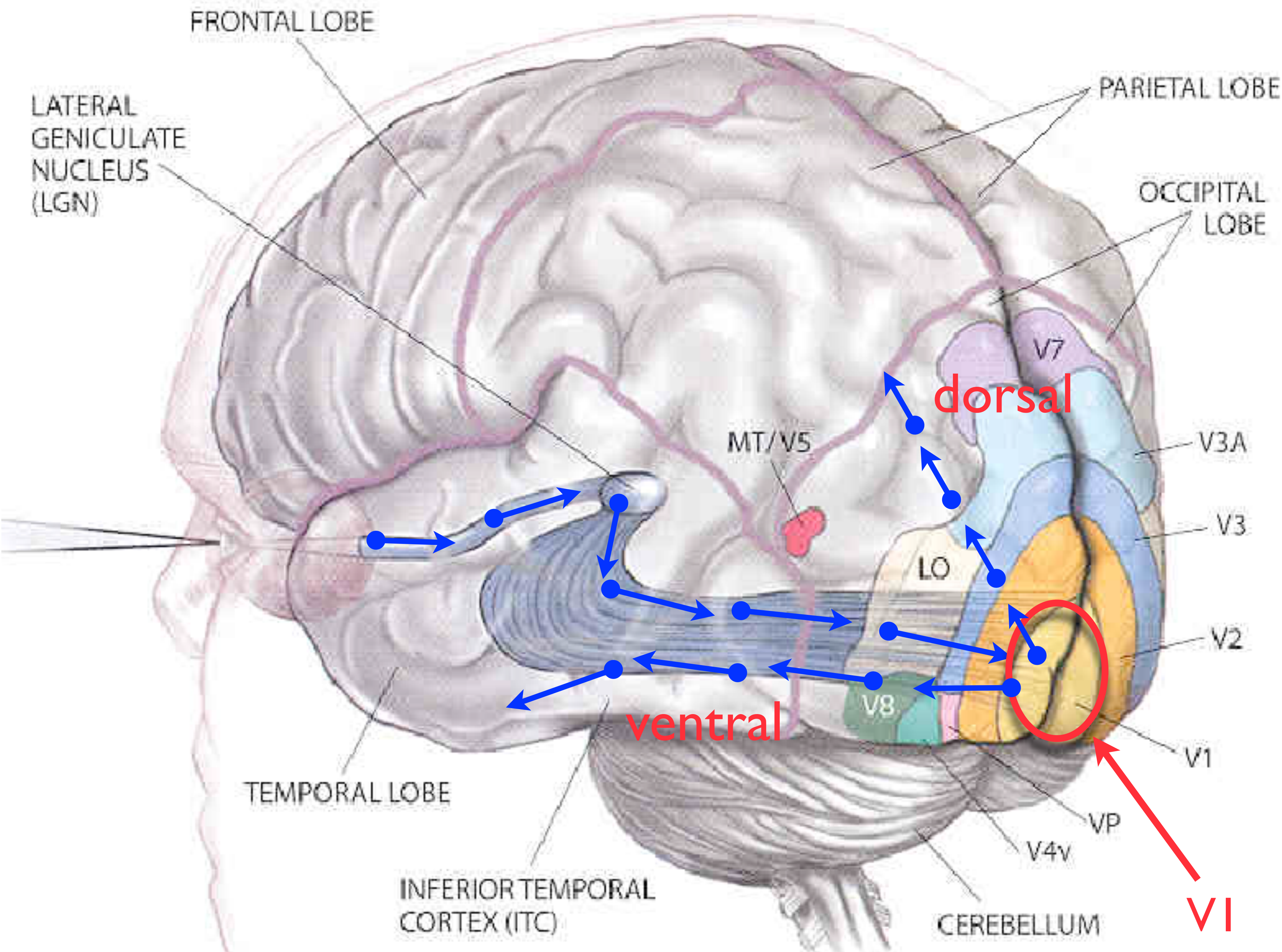
theories of the brain's internal processes of perceptual inference

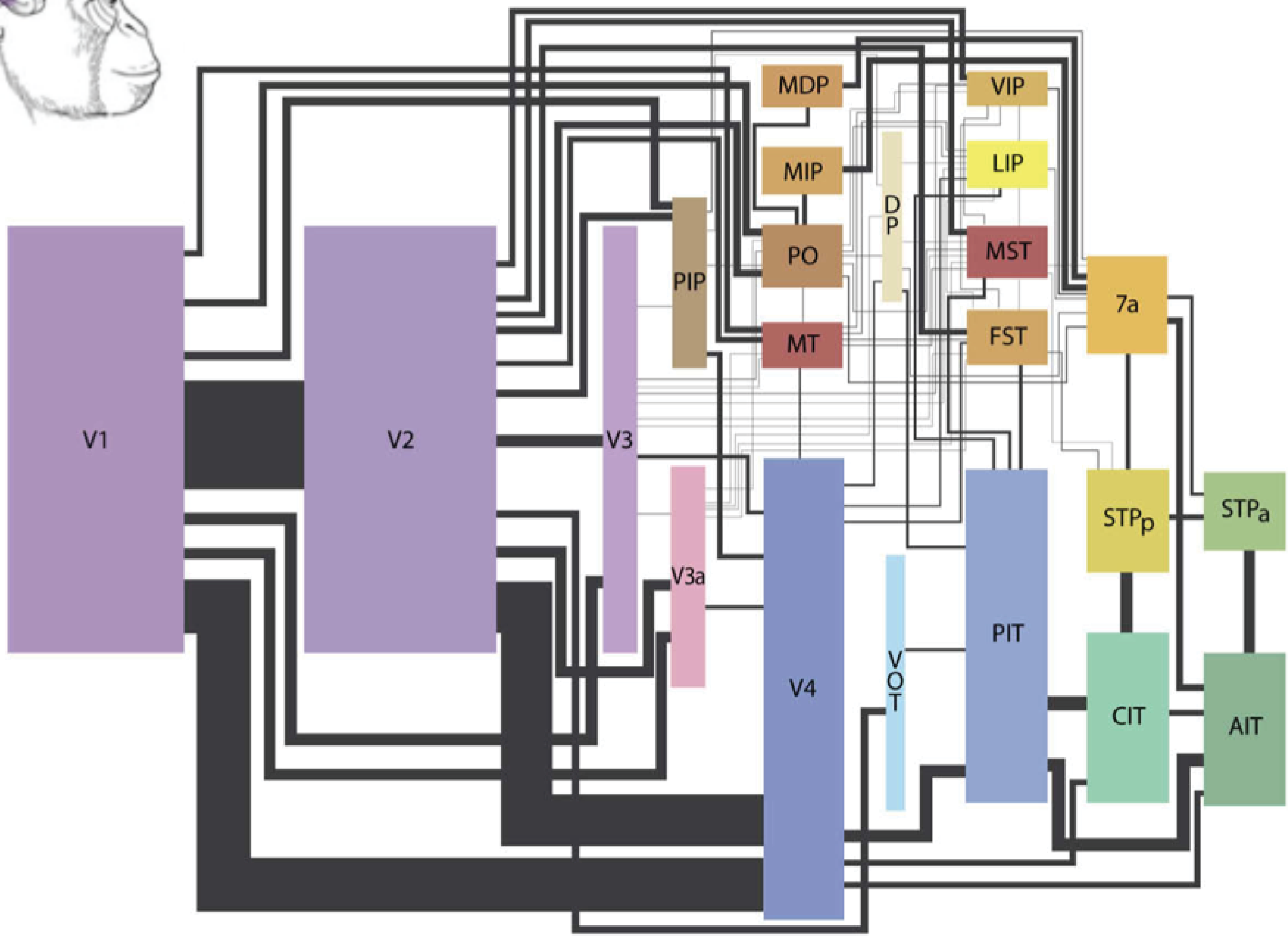
30+ cortical areas that are visually sensitive, often with specific preferences, such as

- localized edges, color,
- motion
- object patches, whole objects,..
- face parts, faces
- bodies,..
- places…



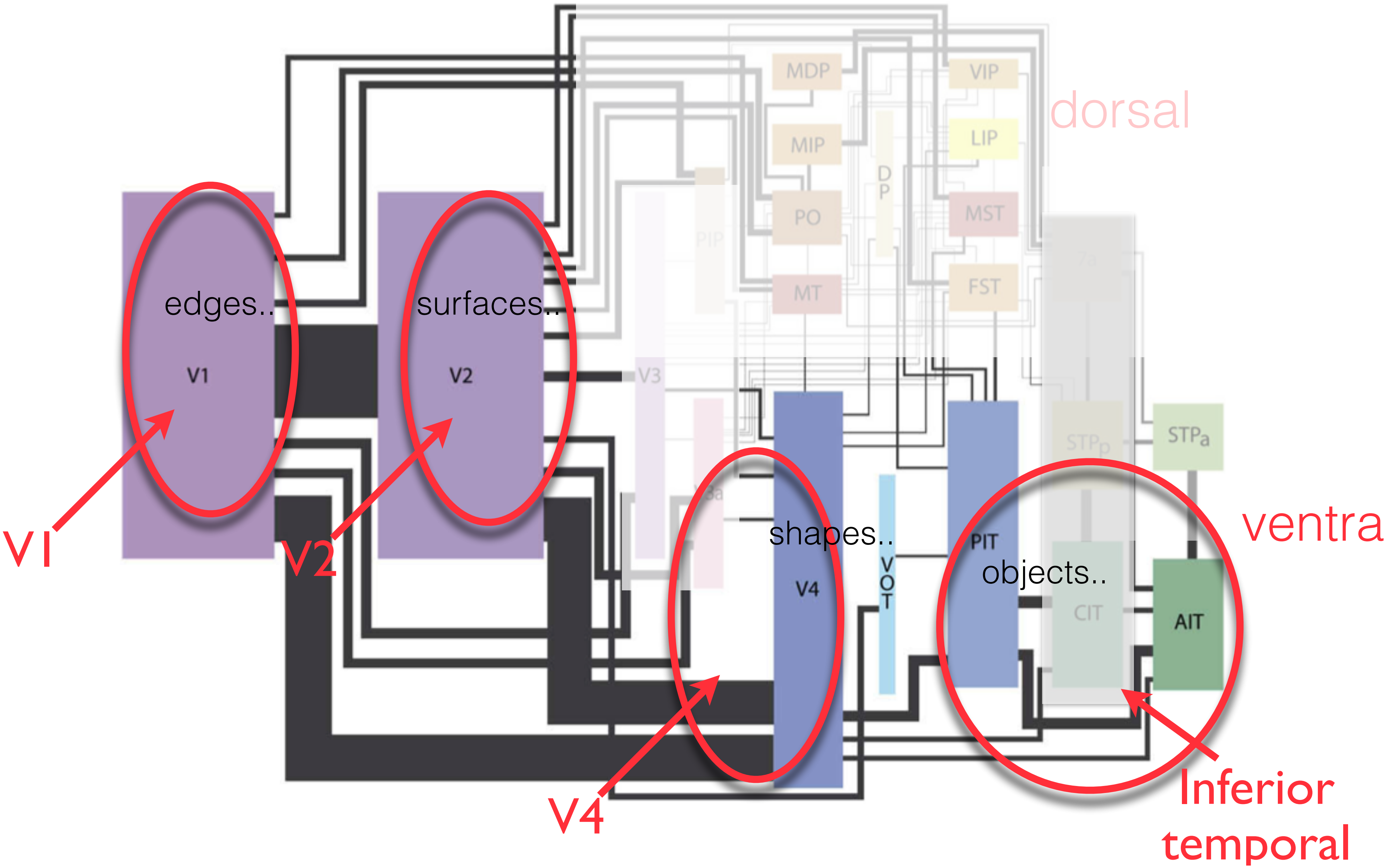
Wallisch, P., & Movshon, J. A. (2008). Structure and Function Come Unglued in the Visual Cortex. *Neuron*, *60*(2), 194–197.


FRONTAL LOBE
LATERAL GENICULATE NUCLEUS (LGN)
PARIETAL LOBE
OCCIPITAL LOBE
V7
dorsal
MT/V5
V3A
V3
LO
V2
V8
ventral
V1
TEMPORAL LOBE
VP
V4v
V1
INFERIOR TEMPORAL CORTEX (ITC)
CEREBELLUM



dorsal

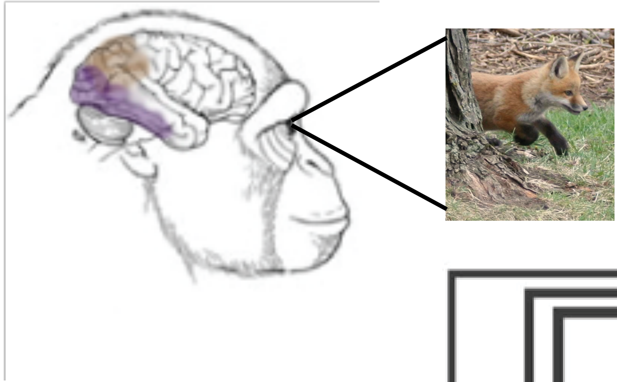
ventral

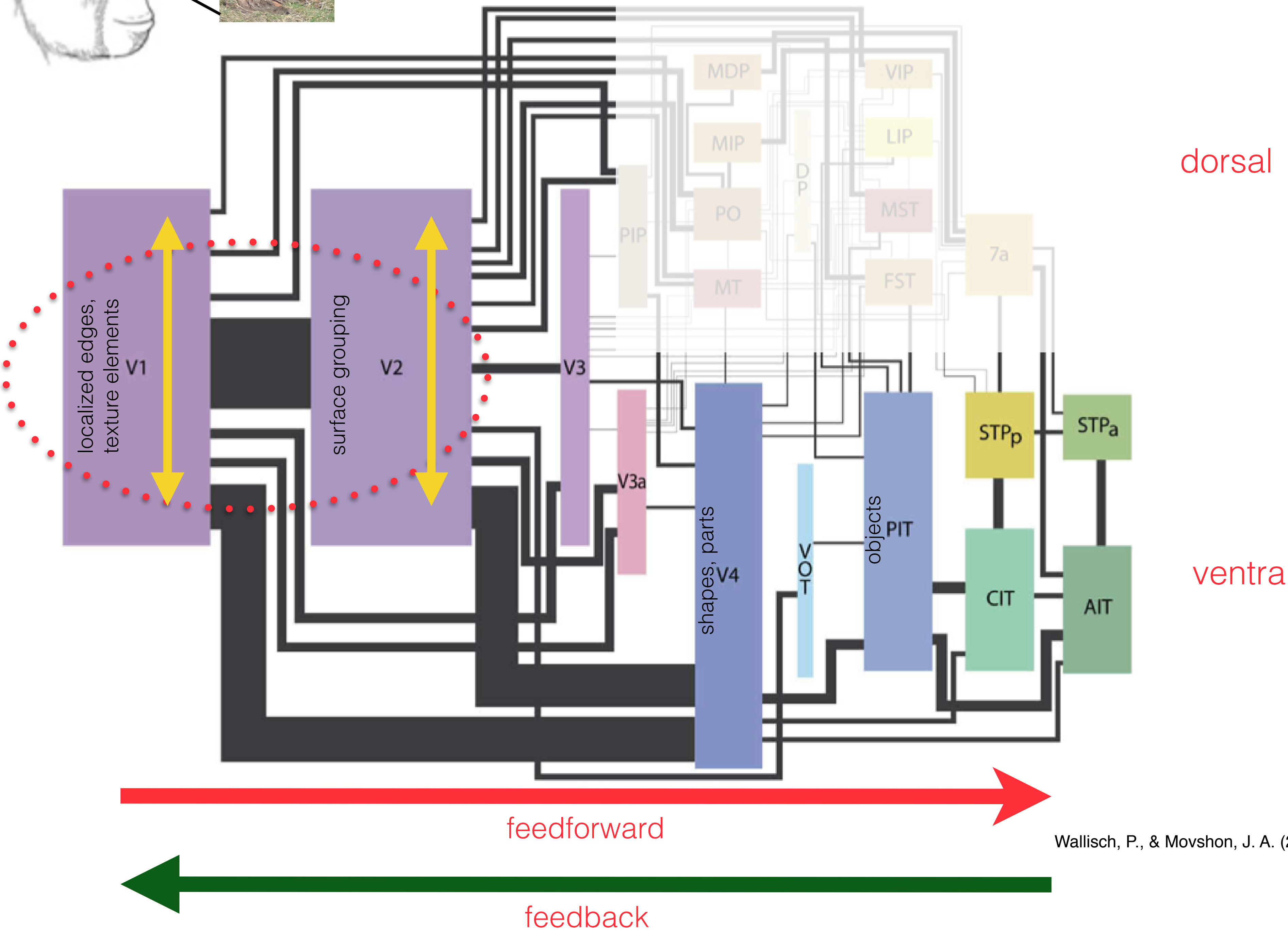


increasing

- receptive field size
- invariance to position, size, illumination, ..
- pattern selectivity

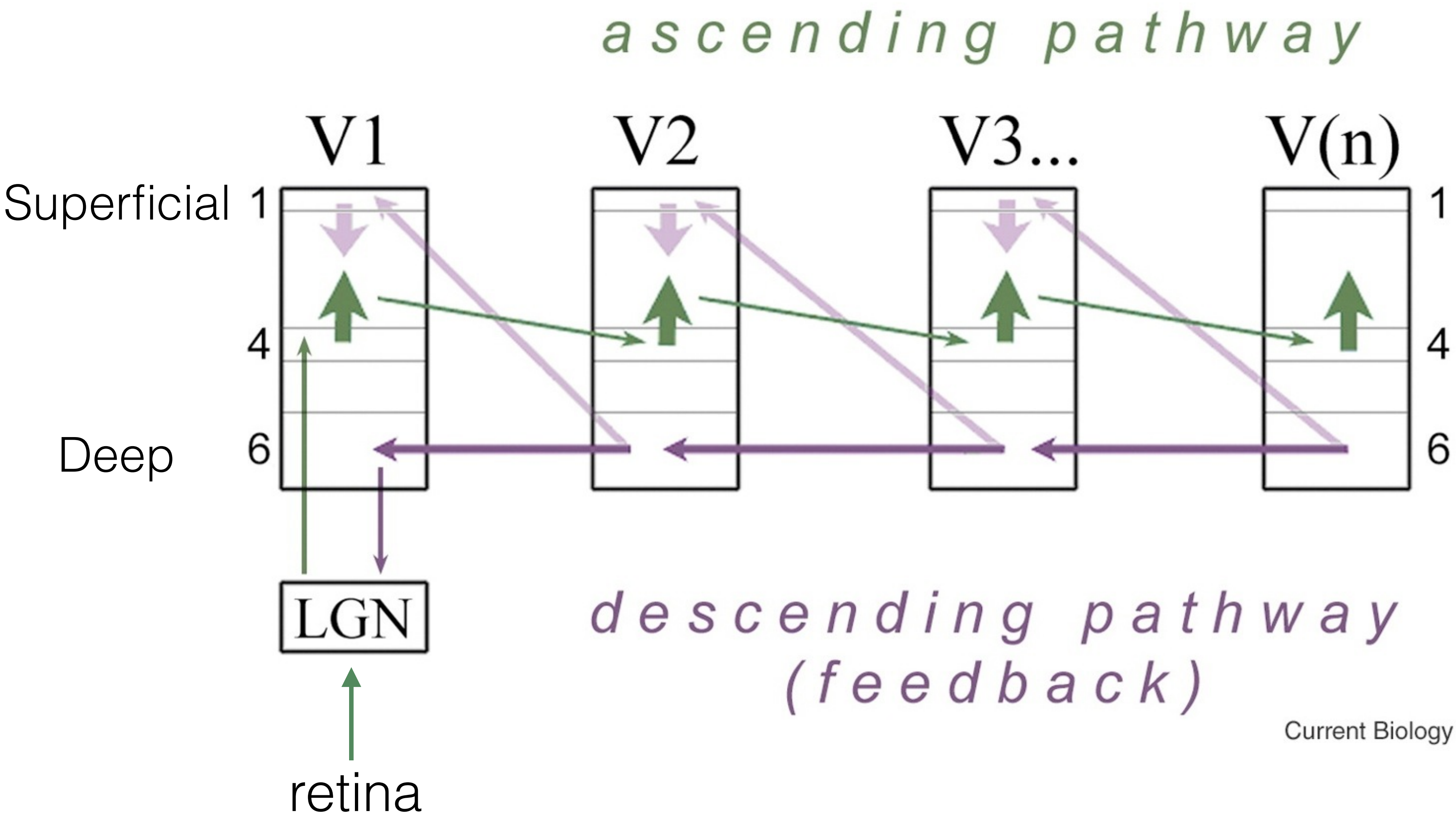
feedforward, lateral and feedback processing



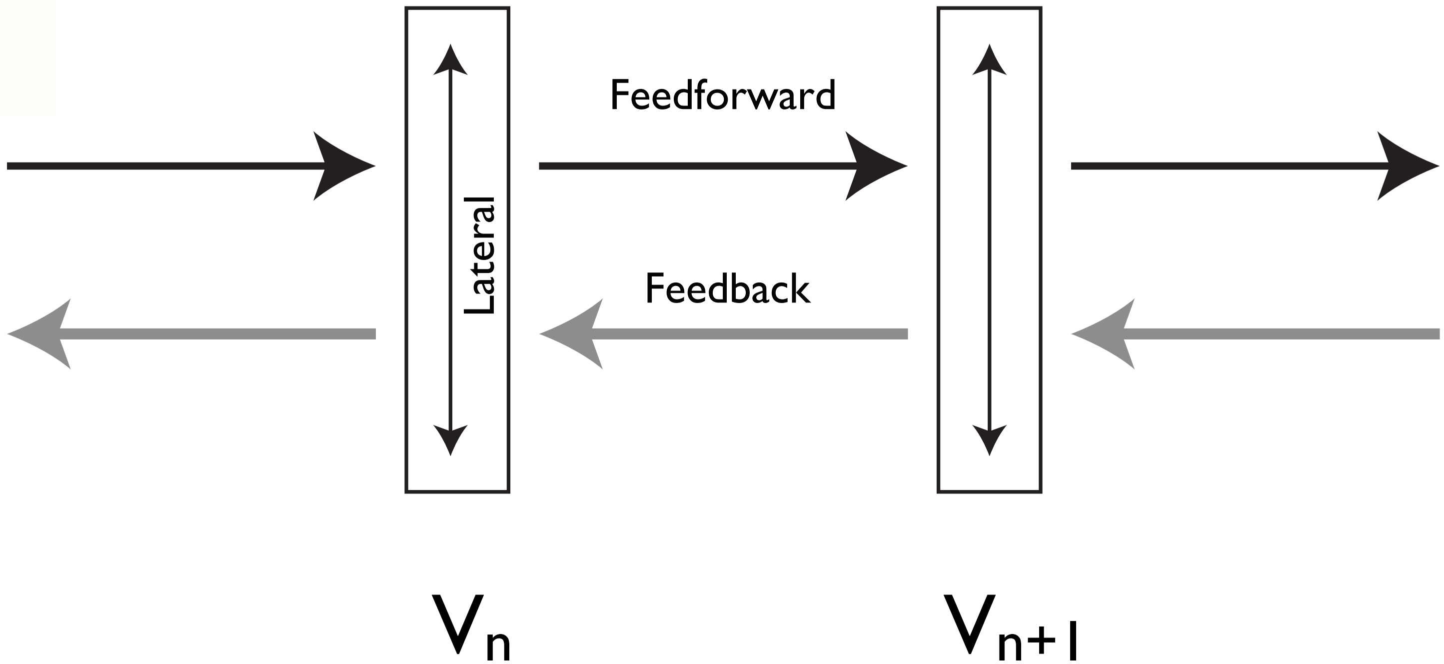


Wallisch, P., & Movshon, J. A. (2008)

Bidirectional processing

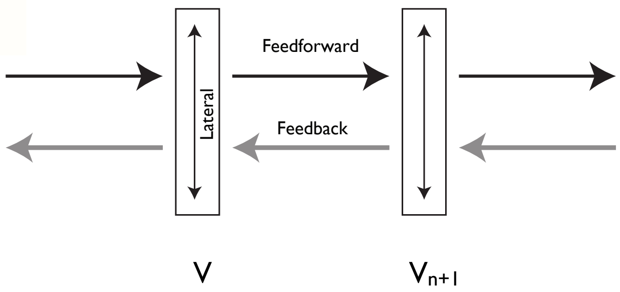


Shipp, S. (2007). Structure and function of the cerebral cortex. CURBIO, 17(12), R443–9. doi:10.1016/j.cub.2007.03.044

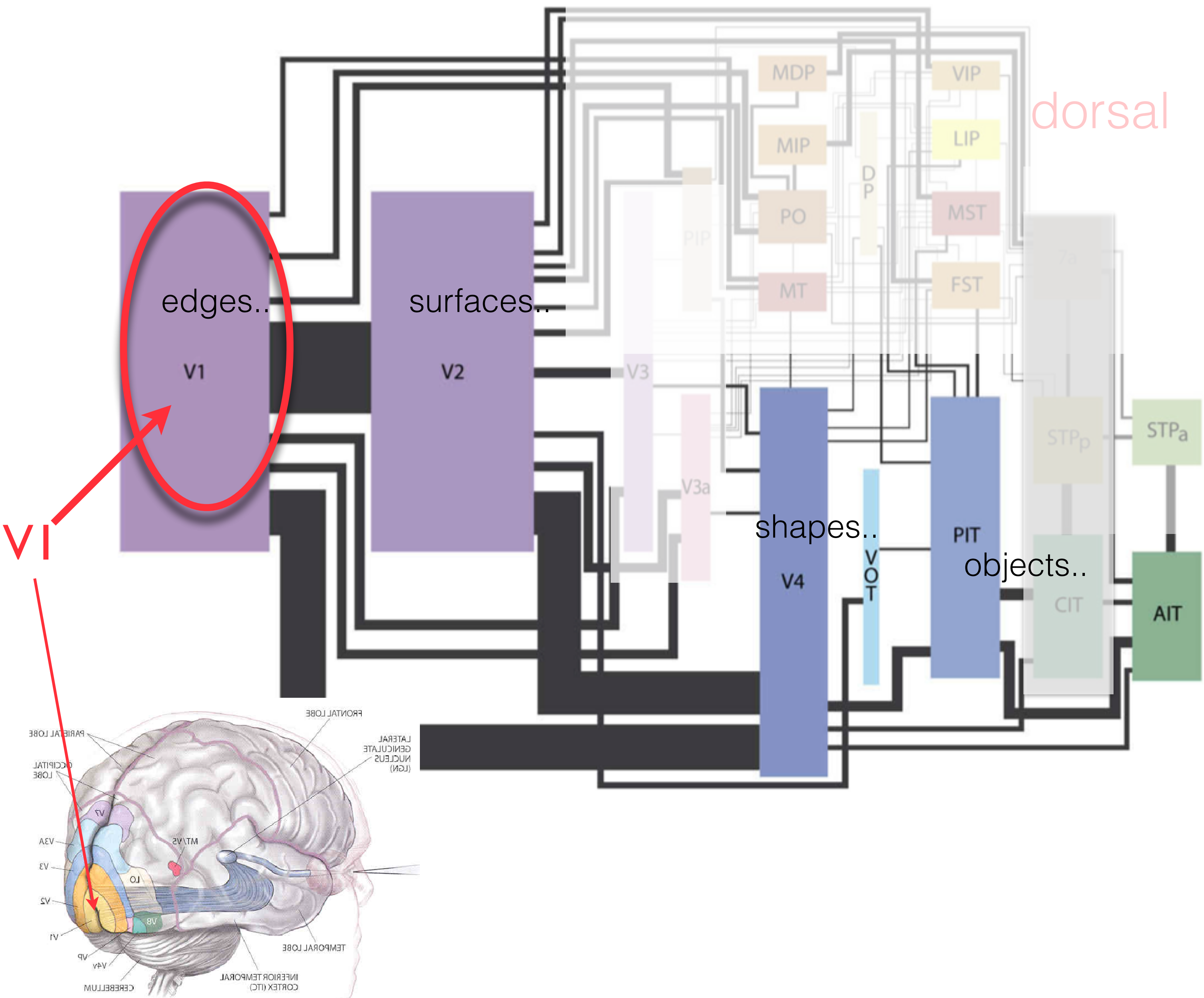


Are there common principles of organization and computation
laterally, feedforward, and feedback?

overview of the rest of the semester



- Lateral organization
  - grouping similar features
  - metabolically and statistically efficient representations
  - neural population codes: representing uncertainty?
- Hierarchical architecture
  - Feedforward functions
  - Feedback functions


edges..
V1
surfaces..
V2
V3
PIP
V3a
MDP
MIP
PO
MT
DP
VIP
LIP
MST
FST
7a
dorsal
STPp
STPa
shapes..
V4
VOT
PIT
objects..
CIT
AIT
V1

primary visual cortex (V1)

local: small hypercolumns consisting of banks of neurons tuned for edge orientation

neurons representing similar features are near on cortical surface

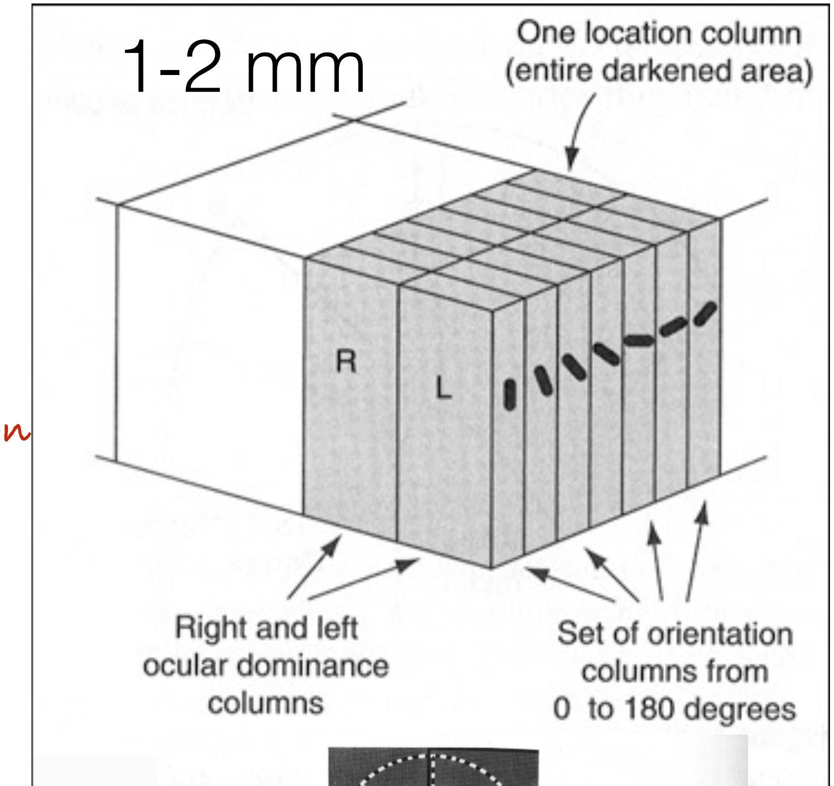
“simple cells” — template matching generic neuron model

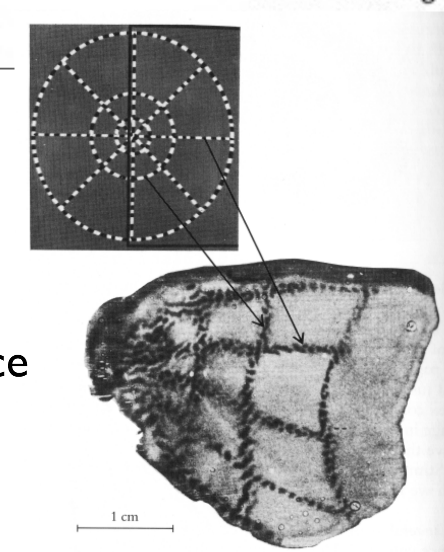
“complex cells” — template matching tolerant to spatial shifts 2 layer generic neuron model

global: hypercolumns arranged retinotopically

neurons receiving information from nearby points in the world are near on cortical surface

neurons receive information laterally, from nearby neurons on cortical surface





lateral organization

Why the organization? The level of abstraction?

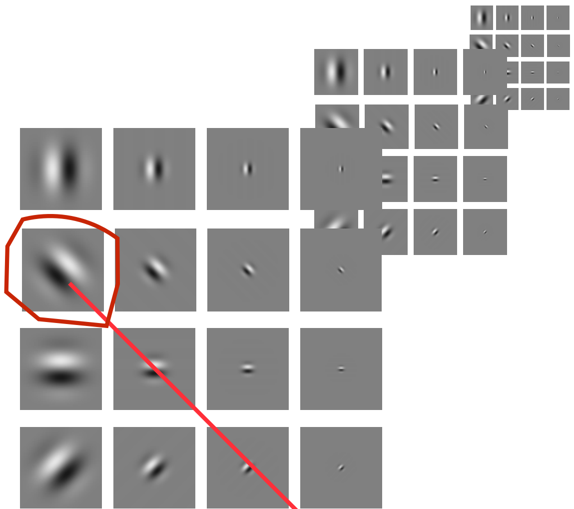
- Keep similar features together for feedforward integration.

will return to this in detail when we get to feedforward functions (and deep, supervised learning)

- Neural population codes
- Lateral computations to group features of similar type—segmentation
- Efficiency constraints
  - Minimum (neural) wiring constraint
  - Efficient representation of sensory input & cost of neural activity
  - Efficient representations for learning

how can receptive field weights be learned (unsupervised)?

unsupervised learning to support supervised learning



standard
non-linear spatial filter
V1 model

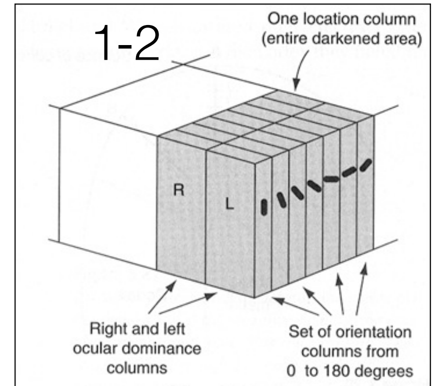
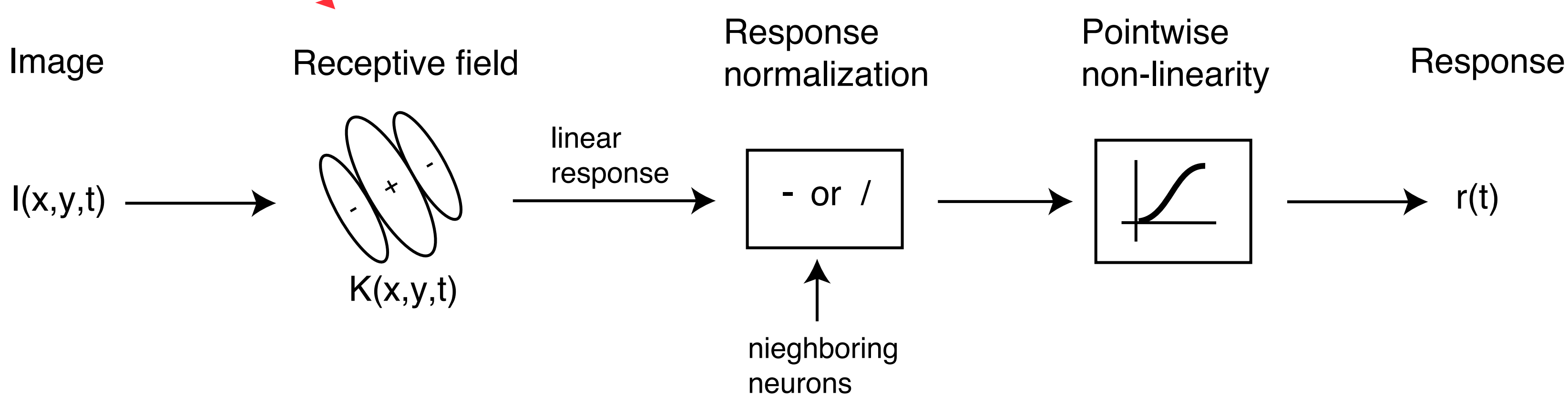
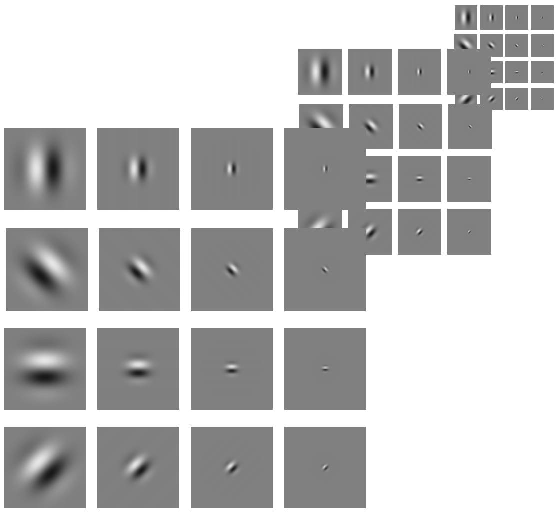


Figure 3.30

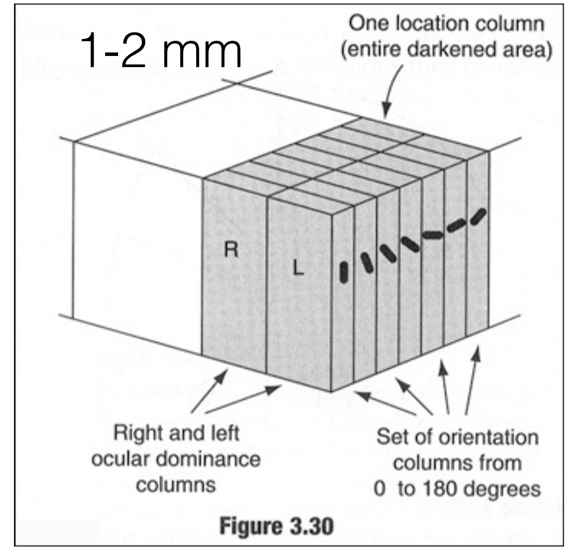
weights



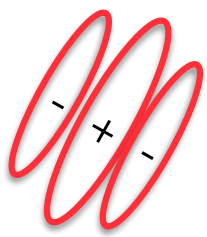
then applied to whole image input via convolution, once for each "channel"

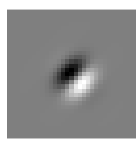


standard non-linear spatial filter V1 model



Receptive field





→

linear

Normalization

→ / →

↑↑↑

Neighboring neurons

just one of many feedforward channels

←··········Sparse coding··········→

to explain orientation/spatial frequency selectivity

lateral organization

Why the organization? The level of abstraction?

- Keep similar features together for feedforward integration.

  will return to this when we get to feedforward functions (and deep, supervised learning)

- Lateral computations to group features of similar type—segmentation

- Efficiency constraints

  - Minimum (neural) wiring constraint

  - Efficient representation of sensory input & cost of neural activity

    how can receptive field weights be learned (unsupervised)?

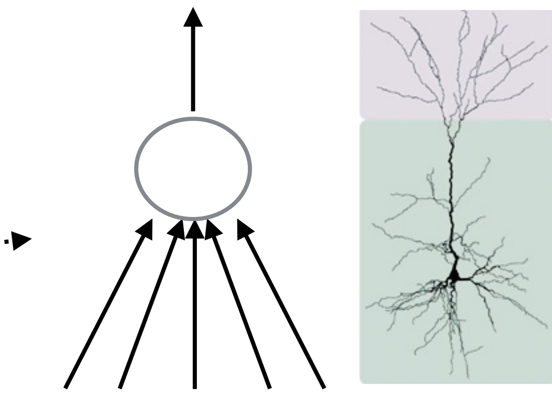
  - Efficient representations for learning

    unsupervised learning to support supervised learning

lateral organization: "maps"

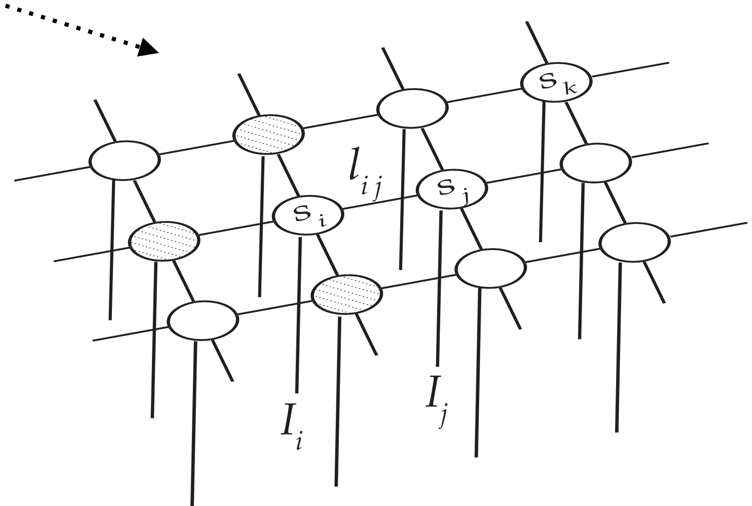
Why the organization? The level of abstraction?

- Keep similar features together for feedforward integration.
- Lateral computations to group features of similar type—segmentation
- Efficiency constraints
  - Minimum wiring constraint
  - Efficient representation of sensory input & cost of neural activity
  - Efficient representations for learning



limited dendritic spread

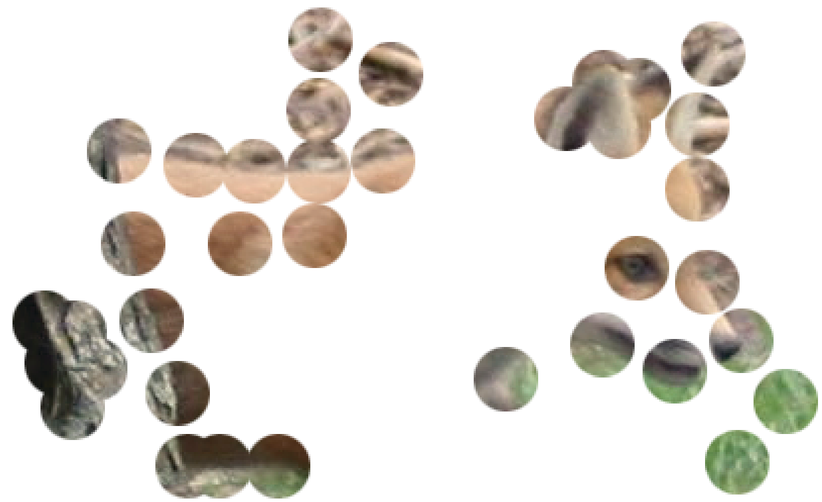
lateral computations?



Markov Random Field models, Gibbs sampling

Grouping

link contours with similar orientations



link regions with similar colors, textures

What should the local features be? How many different types?

V1

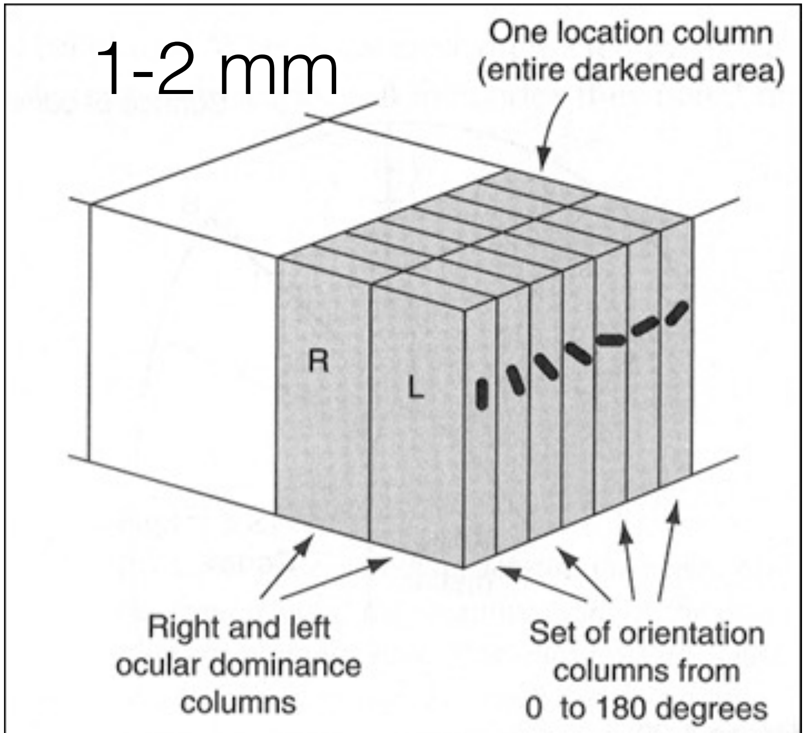
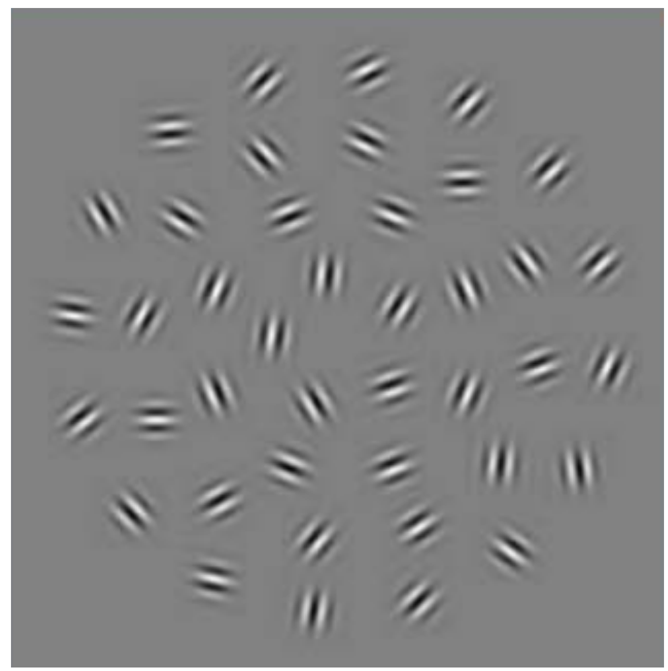


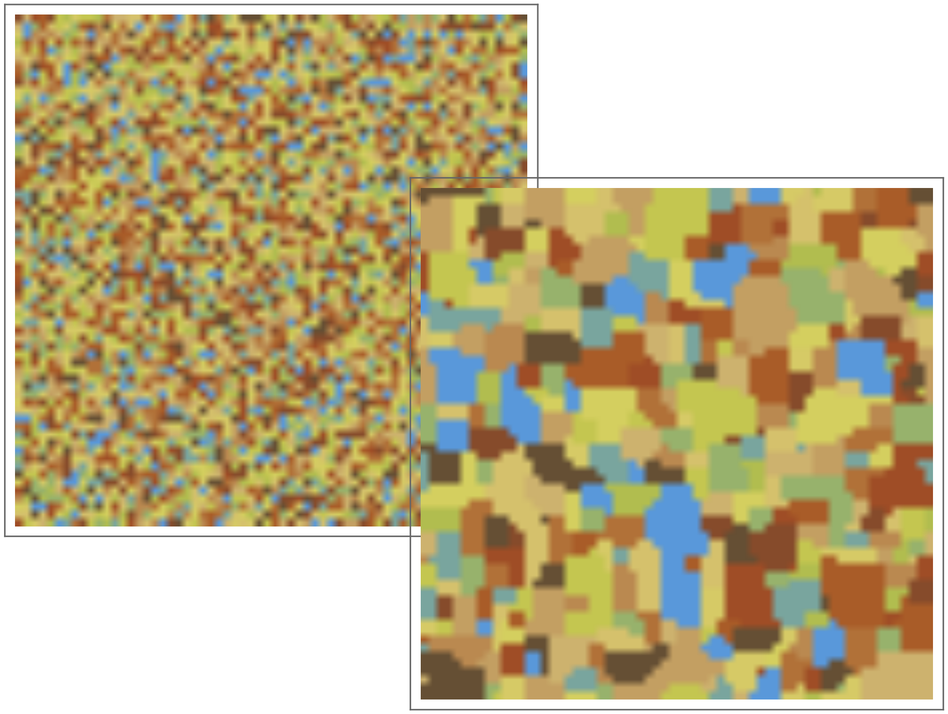
Figure 3.30

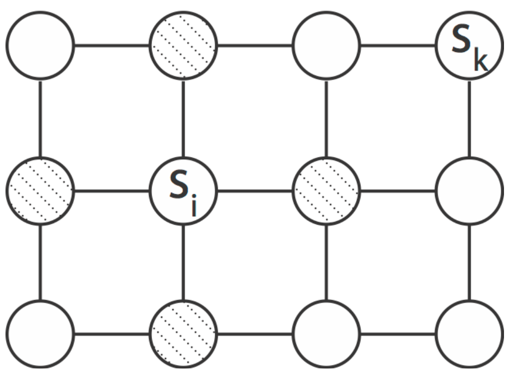
neurons receive information laterally, from nearby neurons on cortical surface, i.e. between hyper columns

The local spatial context may support orientation grouping

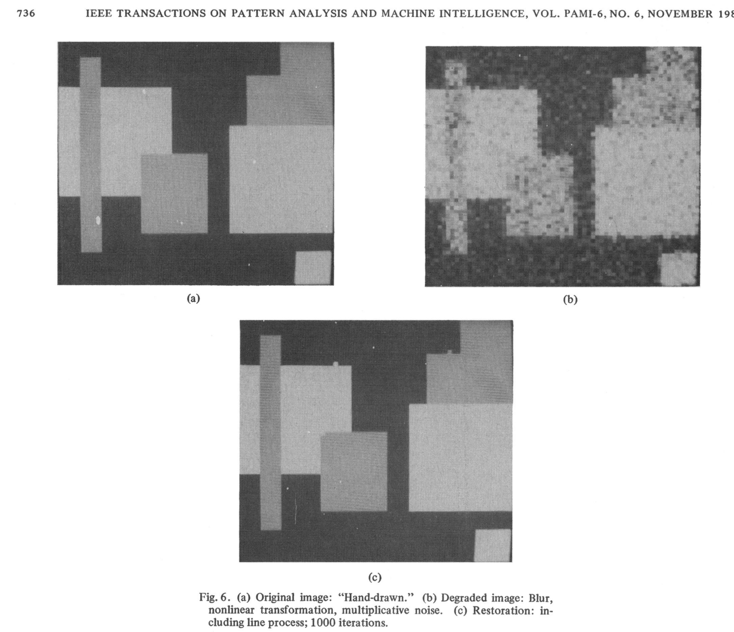


color grouping?



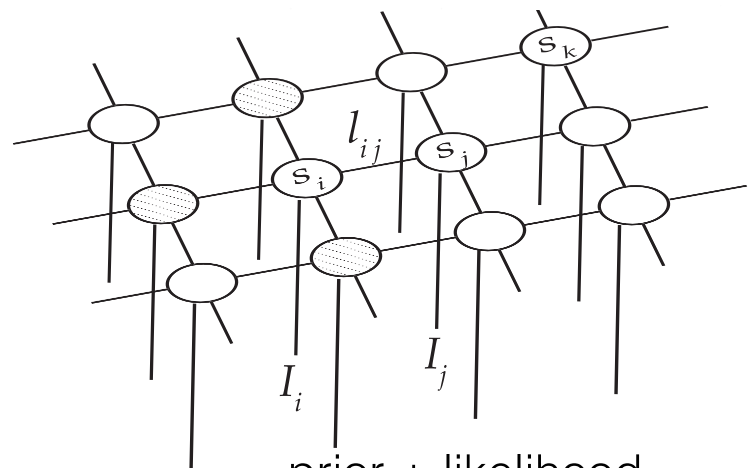


prior


736 IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. PAMI-6, NO. 6, NOVEMBER 1984

(a) (b) (c)

Fig. 6. (a) Original image: "Hand-drawn." (b) Degraded image: Blur, nonlinear transformation, multiplicative noise. (c) Restoration: including line process; 1000 iterations.

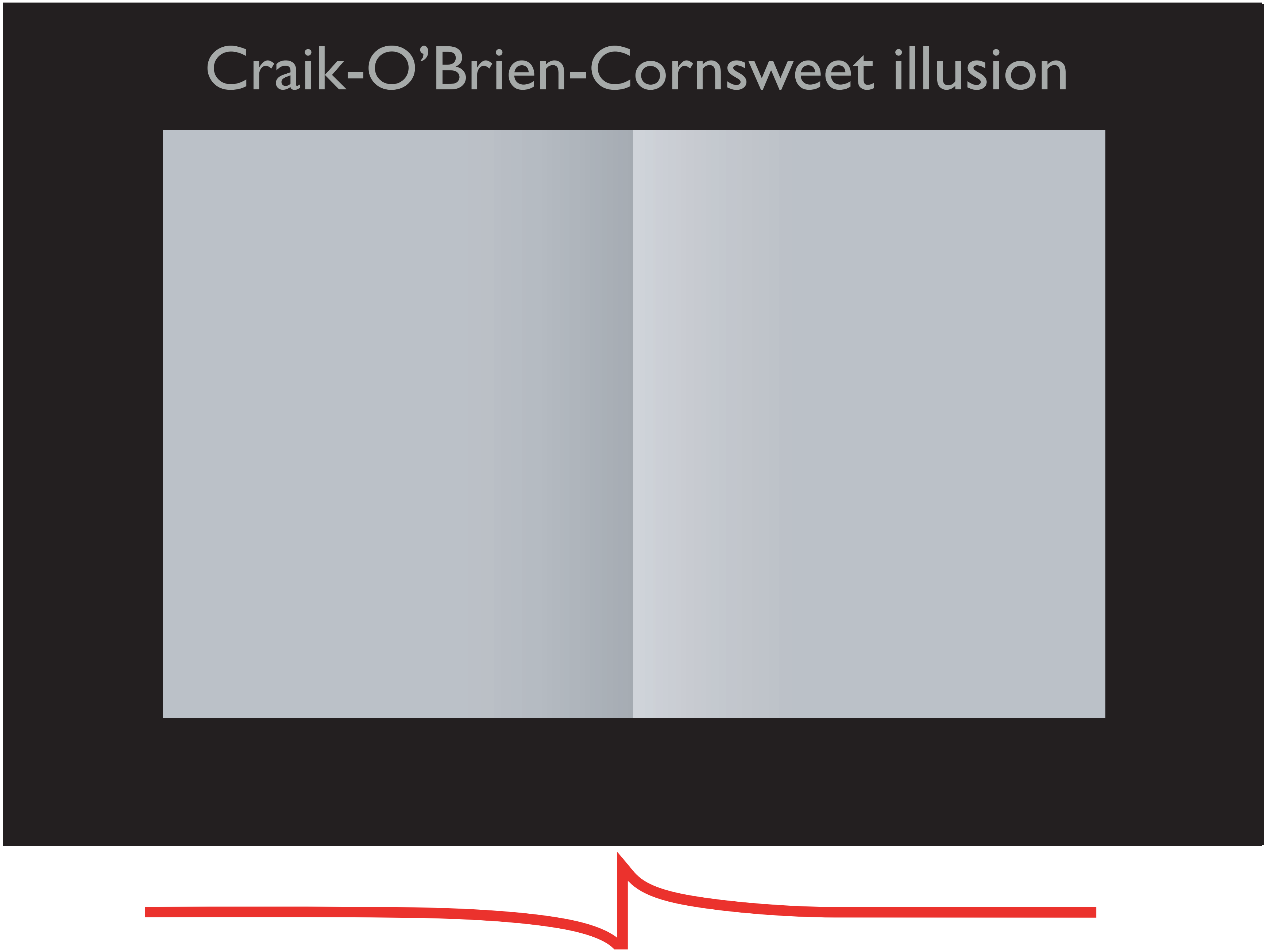


prior + likelihood

..but would the visual system need to "denoise"?

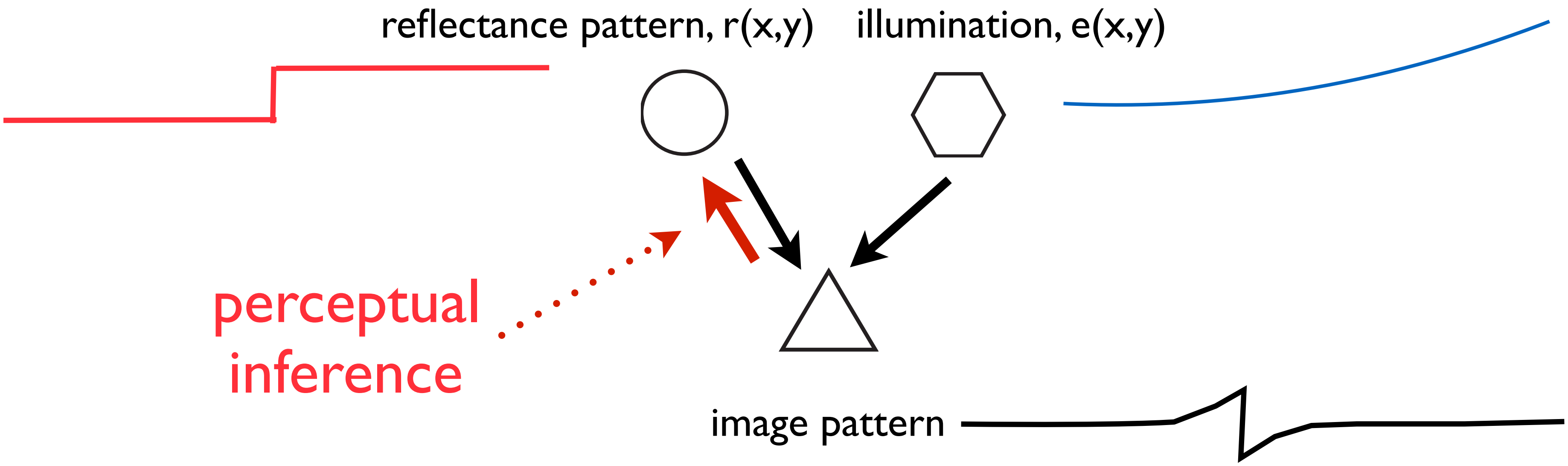
what is noise anyway?

Human fMRI evidence for lateral computations?



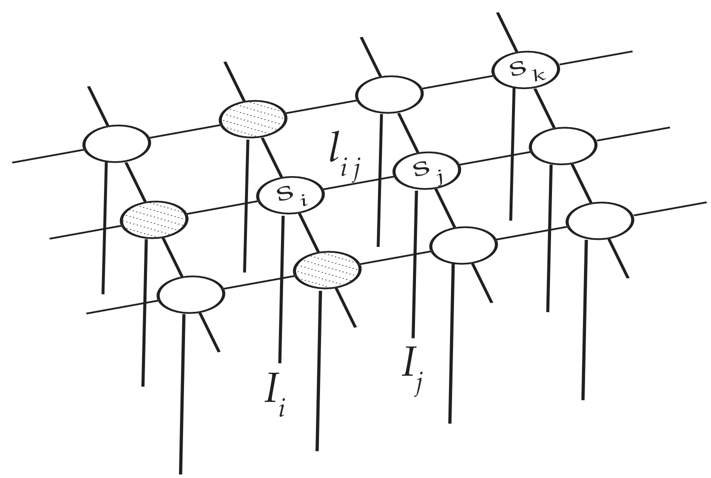
What are the features that are being linked?

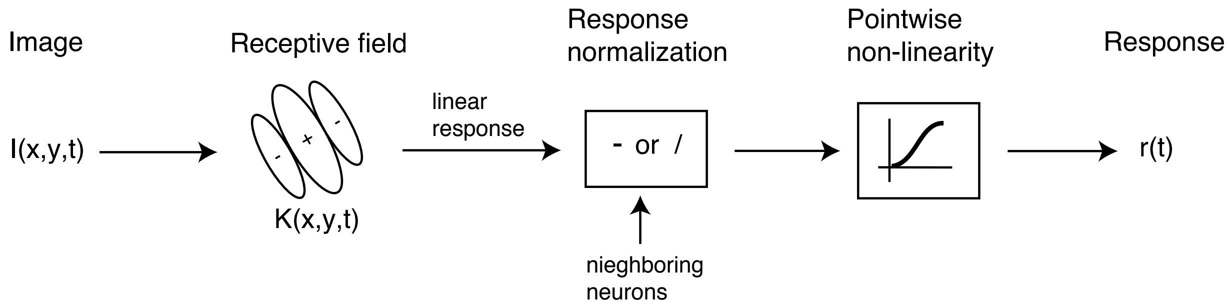
image = f(pigment, illumination) ~ $r(x,y) \times e(x,y)$



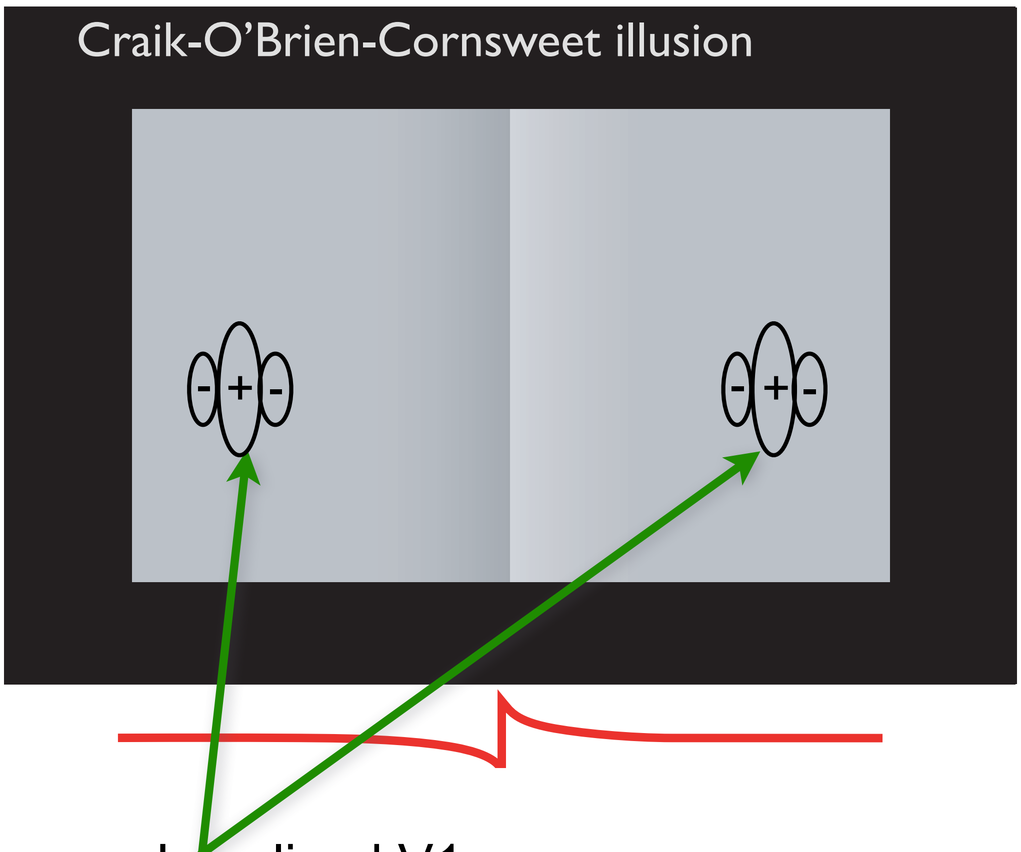
estimate pigment property--the *reflectance*, and discount illumination

prior probabilistic structure:
- illumination spatially smooth
- reflectance is piece-wise constant.
- E.g. gibbs sampler texture demo





Convolution



Localized V1
responses here should be the
same with standard feedforward
model

V1 response follows perceived lightness, not physical intensity

Purely lateral? Don't know. But neuroimaging effect persists with when attention is diverted.

Boyaci, H., Fang, F., Murray, S. O., & Kersten, D. (2007). Current Biology, 17(11), 989–993.

lateral organization

Why the organization? The level of abstraction?

- Keep similar features together for feedforward integration.
- Lateral computations to group features of similar type—segmentation
- Efficiency constraints
  - Minimum wiring constraint
  - Efficient representation of sensory input & cost of neural activity
  - Efficient representations for learning

what constrains receptive field weights?
how to learn?

both unsupervised, and supervised learning methods

unsupervised learning of receptive fields

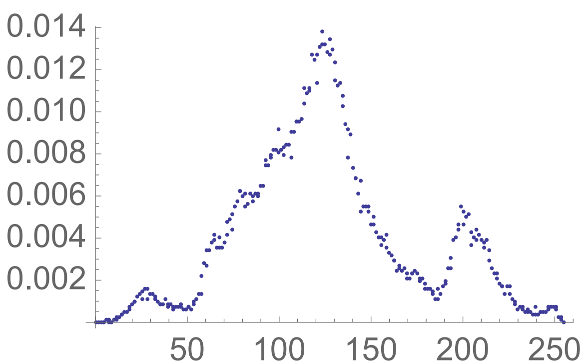
- Unsupervised learning assumes there is statistical structure to be discovered in the sensory input
- Exploit regularities in natural image input to either reduce redundancy or dimensionality, or reduce #active neurons with minimal loss of information.

"efficient coding theories"

Types of structure

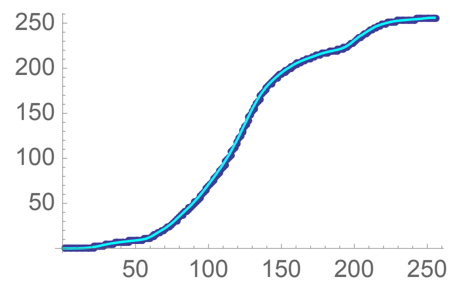
1rst order

What to do with the structure?

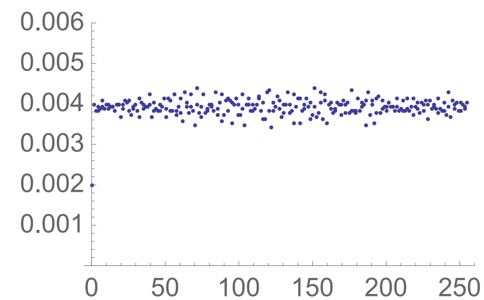


Recode to eliminate it

output

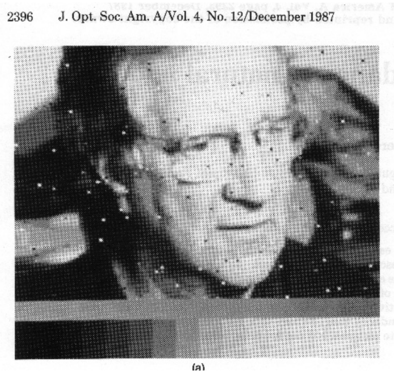


input



single neuron responses are equally probable given natural intensity inputs

the inverse of the CDF method of generating random numbers



Kersten, D. (1987). Predictability and redundancy of natural images. J Opt Soc Am A, 4(12), 2395-2400.

Types of structure

2nd order

Pixel colors can predict the colors of their neighbors

Gives rise to neural network models that are closely related to principles of image compression developed in signal processing theory, as in “difference coding”

R(x) = L(x) - L(x-1)

which exploits the observation that L(x) is often ~ L(x-1)

this looks like lateral inhibition!

$$R(x) = L(x) - \sum_{x' \neq x} w(x - x')L(x')$$

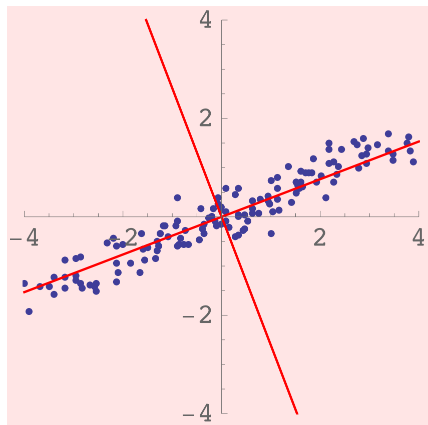
Types of structure

2nd order

Dimensionality reduction via
Principal Components Analysis (PCA) or Singular Value Decomposition (SVD)

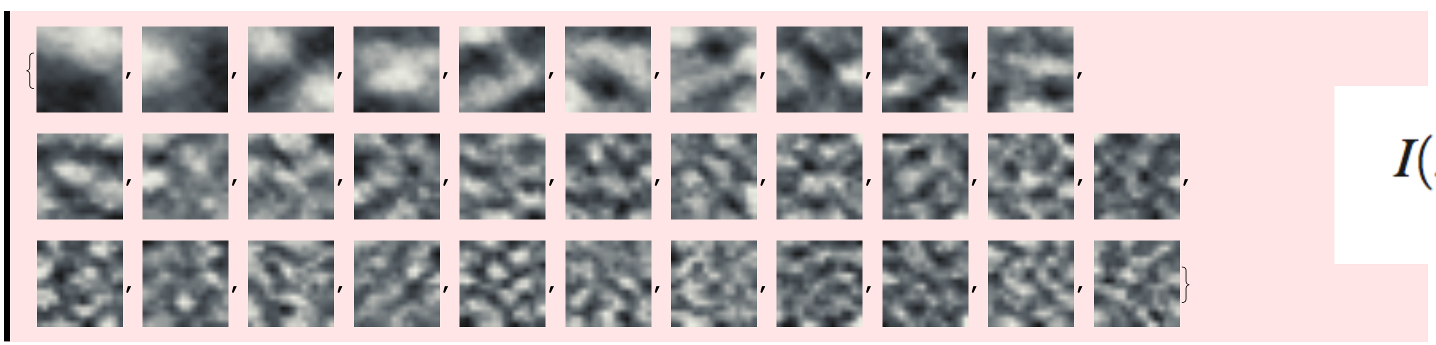
unsupervised learning

2 pixel example:



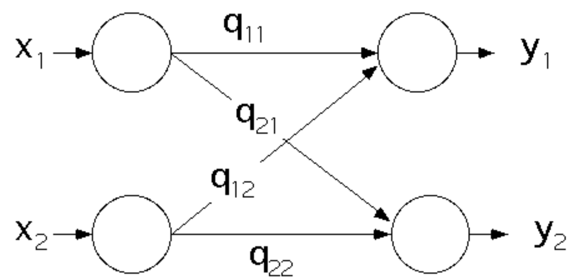
decorrelates the input

and provides the basis for throwing out dimensions



$$I(x,y) = \sum_{i=1}^{n} A_i(x,y) s_i$$

Principal Components Analysis (PCA) with neural networks



Hebbian learning + Oja's rule to normalize weights:

$$\Delta q_{ij} = \alpha\left(x_j y_i - q_{ij} y_i^2\right)$$

Oja's rule automatically normalizes:

$$\sum_{i,j} q_{ij}^2 = 1$$

…but because of symmetry, this network will only pull out the first principal component, and does it twice (in this case)

A solution?

$$\Delta q_{ij} = \alpha\left(x_j y_i - y_i \sum_{k=1}^{i} q_{kj} y_k \right)$$

Sanger, T. (1989). Optimal unsupervised learning in a single-layer linear feedforward neural network. Neural Networks, 2, 459-473.

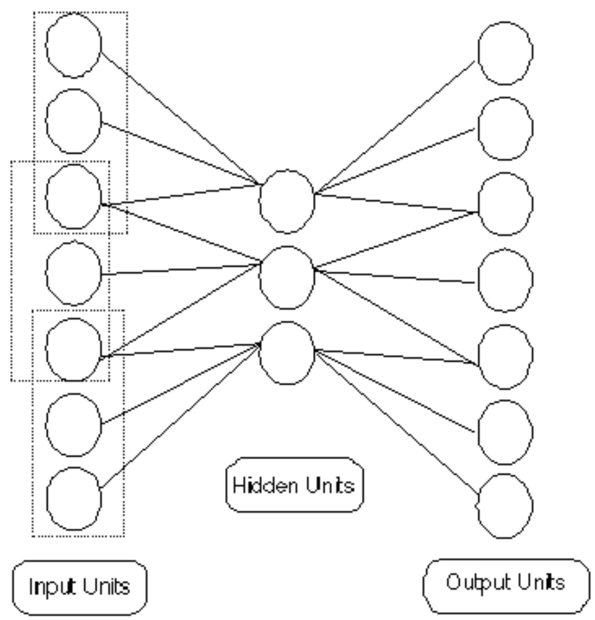
…but this still seems dissatisfying because one neuron would do lots of work, the next less so, and the next even less, etc..

A solution?

"autoencoder networks"

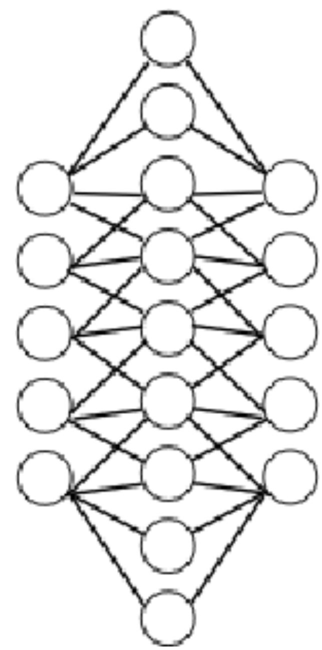
use backprop to find weights that encourage L to predict its own values: input L close to the output L' :

training pairs: $\{L_i, L_i\}$



$L \sim L'$

finds subspace that captures larger fraction of the variance



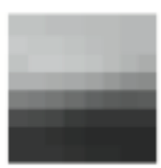
reduce or expand dimensionality

In PCA, the number of basis functions or vectors is less than or equal to the dimensionality of the input

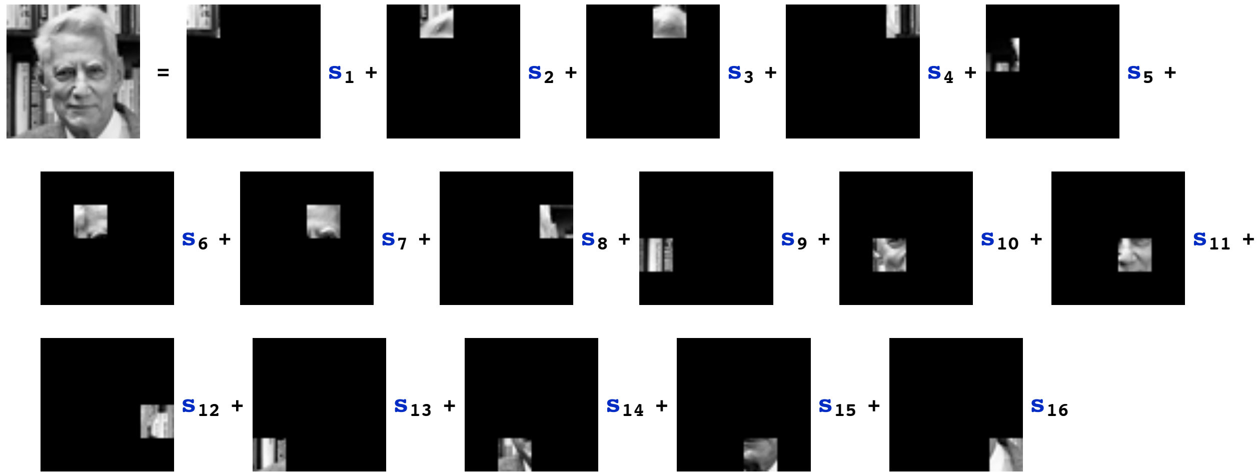
But what if “efficiency” has another meaning, e.g. represent a high-dimensional input (an image) with as few features as possible?

…and we allow for over-complete representations where the number of feature detectors could be more than the dimensionality of the input

$$I(x,y) = \sum_{i=1}^{n} A_i(x,y)s_i \qquad\qquad s_i = \sum_{x,y} W_i(x,y)I(x,y).$$

 $=\ s_1 \cdot$  $+\ s_2 \cdot$  $+ \cdots +\ s_k \cdot$ 

(see Lecture 5)



only a few features required for one image…but what if we wanted to have a set of features, or "dictionary" that was in "good" for all natural images?

Good, efficient representation is interpreted as finding the receptive field weights that minimize the sum of squared errors AND # active neurons

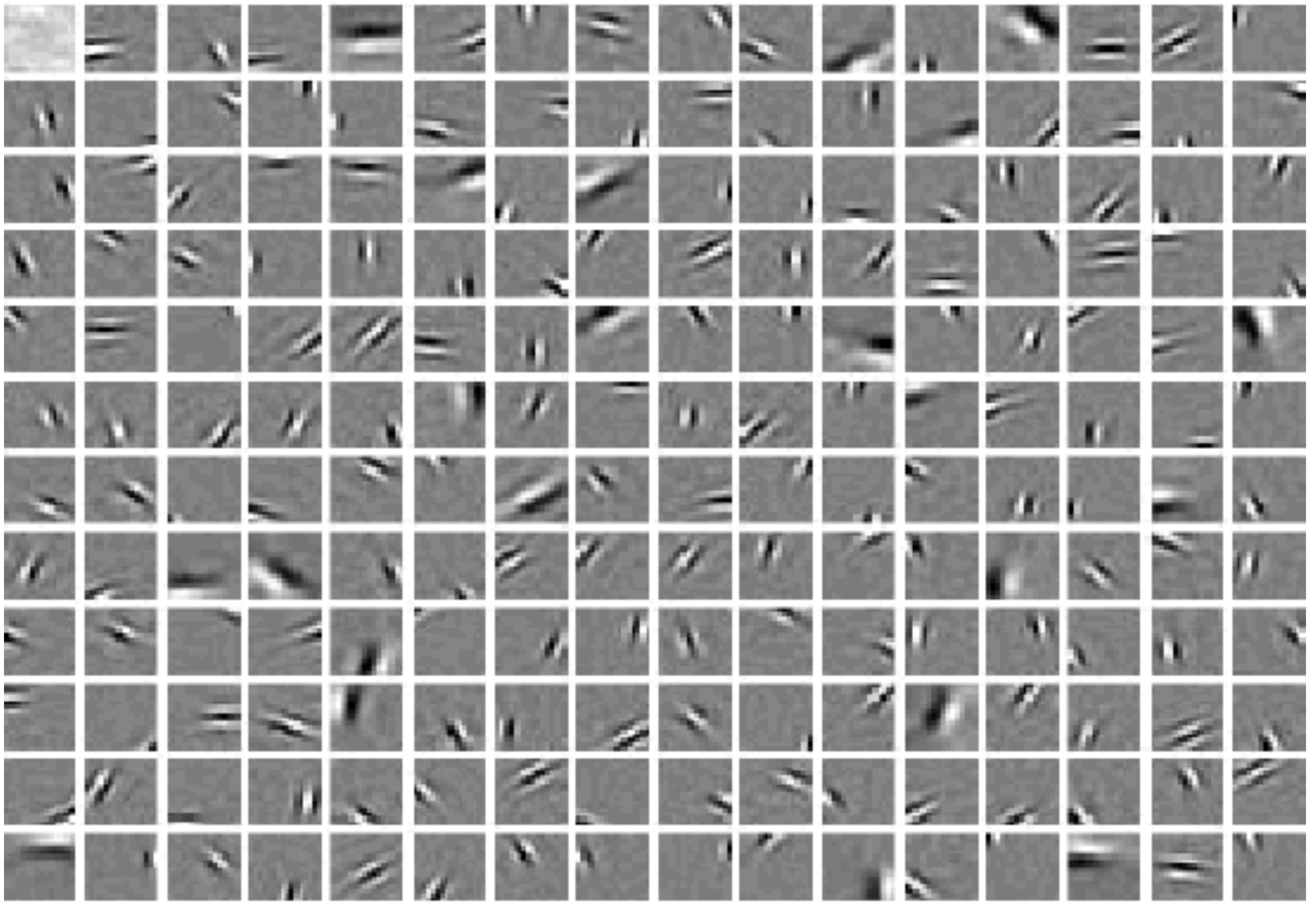
so given L(x,y) in a set of images find the $A_i(x,y)$'s that minimize:

$$[L(x,y) - \sum_i s_i A_i(x,y)]^2 + \sum_i B(s_i)$$

penalizes loss of information about the image

penalizes too many active neurons

the $A_i(x,y)$'s



"sparse coding"

Olshausen & Field's model of V1 receptive fields

captures localized sensitivities to orientation and spatial frequency

Higher-order structure?

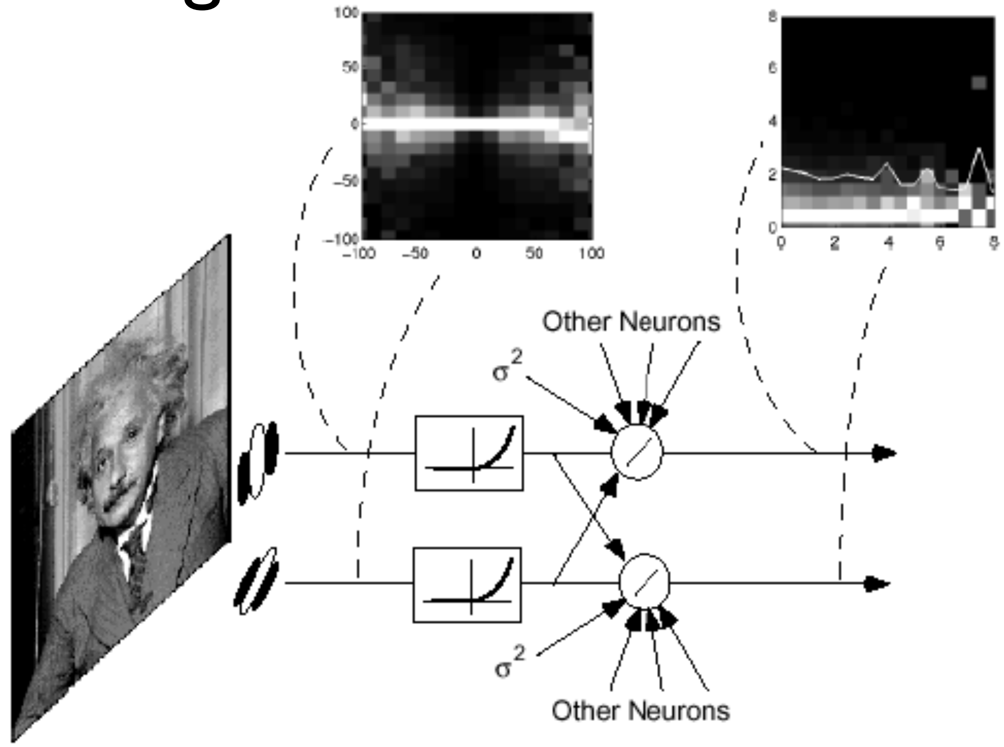


Figure 1: Illustration of image statistics as seen through two neighboring receptive fields. Left image: Joint conditional histogram of two linear coefficients. Pixel intensity corresponds to frequency of occurrence of a given pair of values, except that each column has been independently rescaled to fill the full intensity range. Right image: Joint histogram of divisively normalized coefficients (see text).

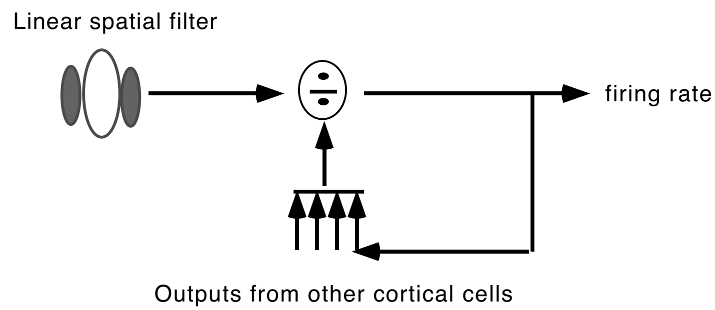
responses of linear model neurons with receptive fields that are close in space, preferred orientation or spatial frequency are not statistically independent

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. Nature Neuroscience, 4(8), 819–825.

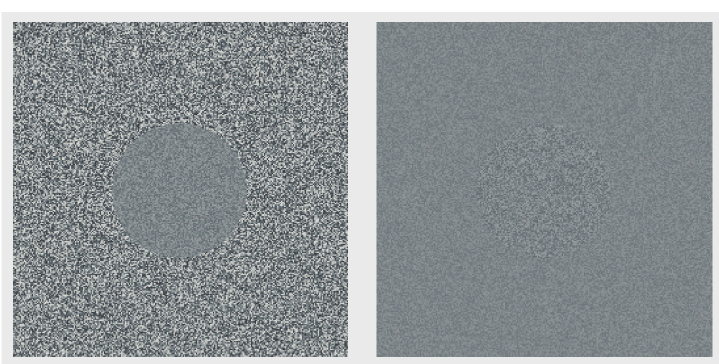
Higher-order structure?

Accounts for neurophysiological responses of neurons in V1.

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. Nature Neuroscience, 4(8), 819–825.



$$R_i = \sigma\left(\sum_{j=1}^{n} w_{ij} L_j\right) \Big/ \sum_{k \in N_i} R_k^2$$



From Heeger

The middle disks have the same physical luminance variance, but the one on the right appears more “contrasty”, i.e. to have higher variance.

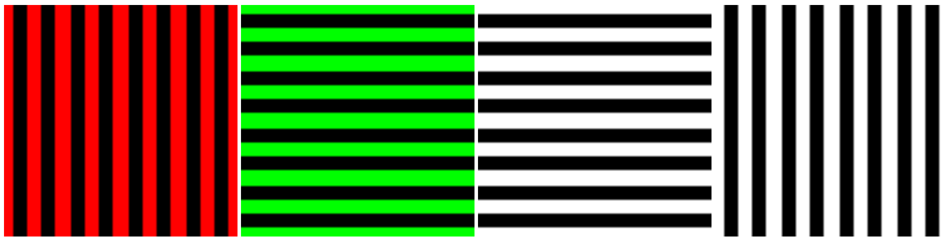
This may be a behavioral consequence of an underlying non-linearity in the spatial filtering properties of V1 neurons involving “divisive normalization” derived from measures of the activity of other nearby neurons.

More on decorrelation:

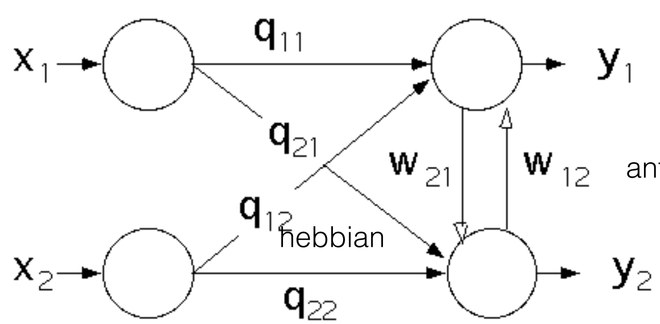
contingent adaptation

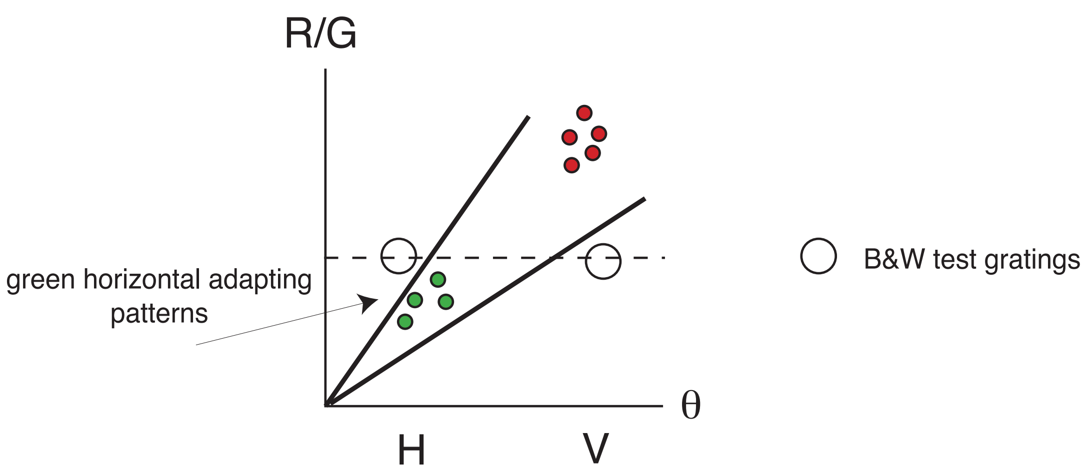
Contingent Adaptation: McCollough effect

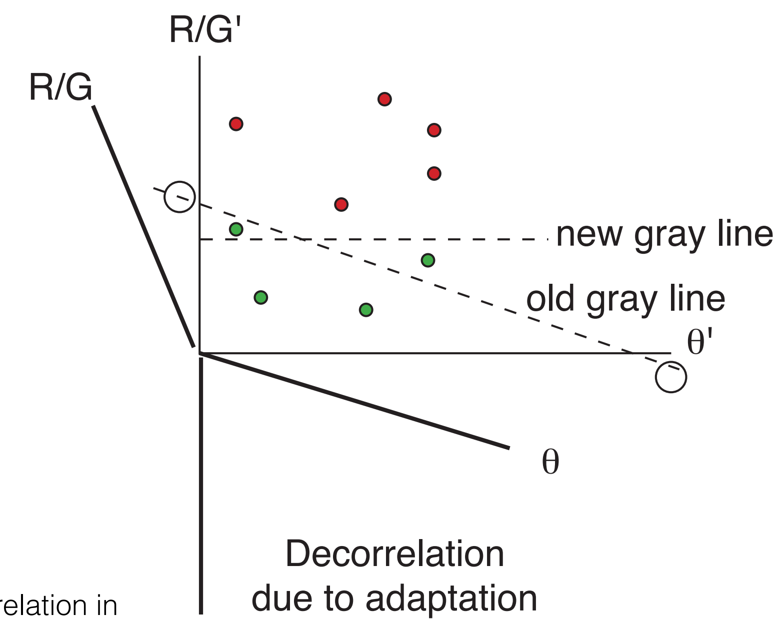
Out[215]=



McCollough, C. (1965, 3 September 1965). Color Adaptation of Edge-Detectors in the Human Visual System. Science, 149, 1115-1116.







Decorrelation due to adaptation

Barlow, H. B., & Foldiak, P. (1989). Adaptation and decorrelation in the cortex. In C. Miall, R. M. Durban, & G. J. Mitchison (Ed.), The Computing Neuron Addison-Wesley.

lateral organization

Why the organization? The level of abstraction?

- Keep similar features together for feedforward integration.
- Lateral computations to group features of similar type—segmentation
- Efficiency constraints
    - Efficient representation of sensory input & cost of neural activity
    - Efficient representations for learning
    - Minimum wiring constraint

to keep similar features near..
but V1 is ~ 2D, and many features!

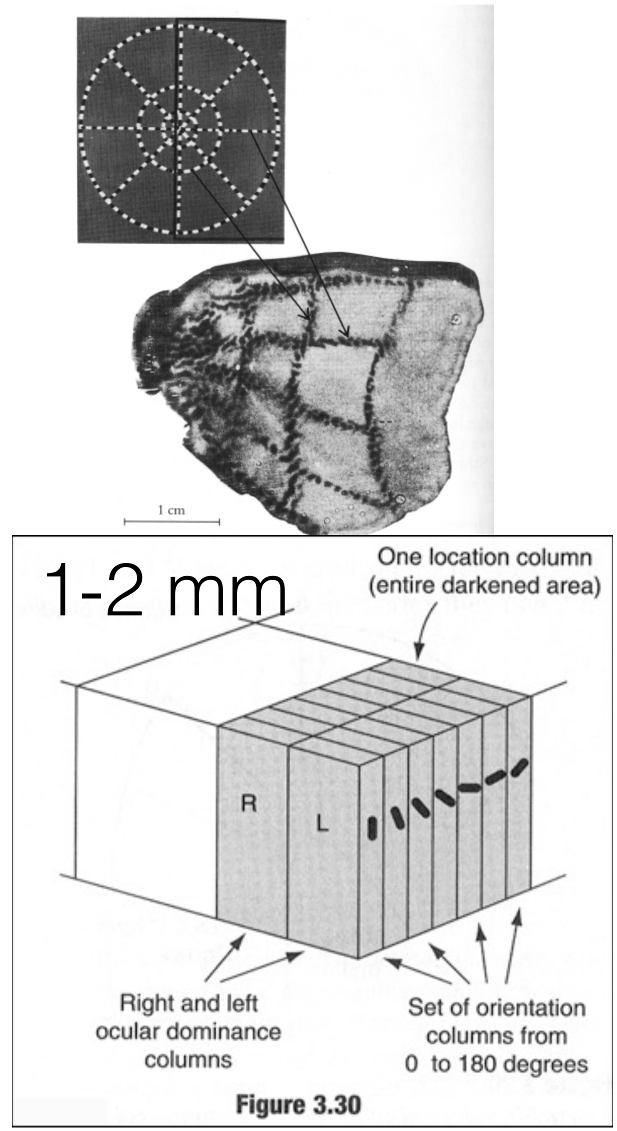
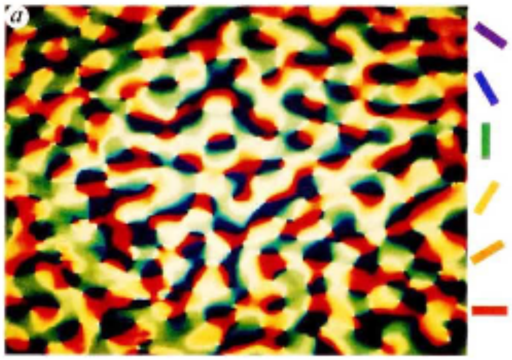


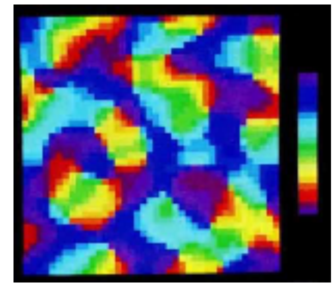
Figure 3.30

how can layout be learned?

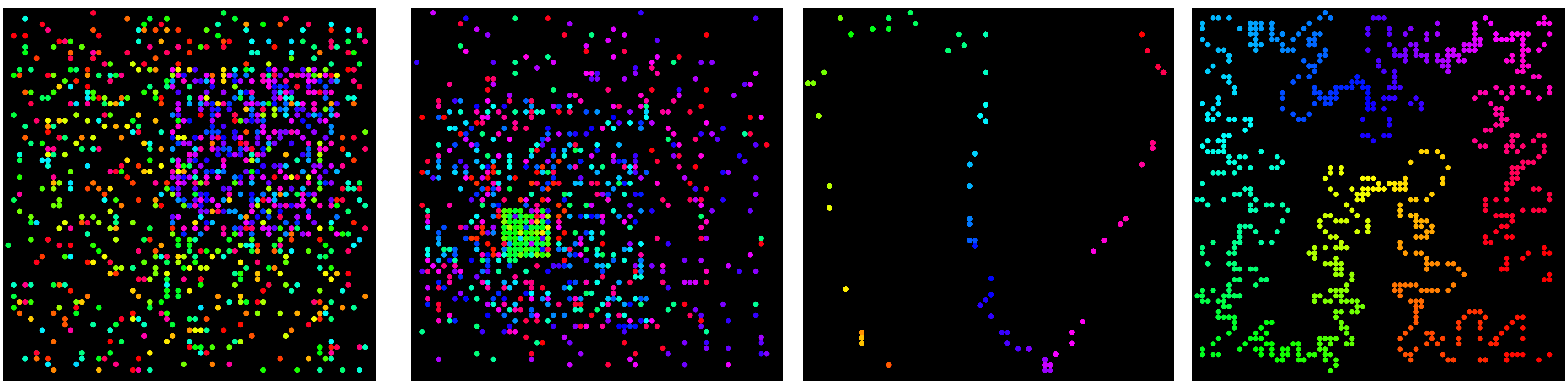
minimum wiring



Ts'o, D. Y., Frostig, R. D., Lieke, E. E., & Grinvald, A. (1990, 27 July 1990). Functional Organization of Primate Visual Cortex Revealed by High Resolution Optical Imaging. Science, 249, 417-420.



Durbin, R., & Mitchison, G. (1990). A dimension reduction framework for understanding cortical maps. Nature, 343, 644-647.



Kohonen map demo: Mapping 2D to 1D

Just V1?

Tanaka, K. (2003). Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities. Cereb Cortex, 13(1), 90-99.

next lectures

- Unsupervised organization of features
  - PCA, SVD
  - non-orthogonal mappings and contingent adaptation
    - anti-hebbian learning
    - auto encoders

scientific writing too

- Relation to machine learning: clustering, EM, K-means
- Minimum wiring algorithm (Kohonen) for global cortical organization
- Neural population codes for representing probabilistic information

next lectures

- Hierarchical processing in depth
  - feedforward—deep convolutional networks
  - feedback