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

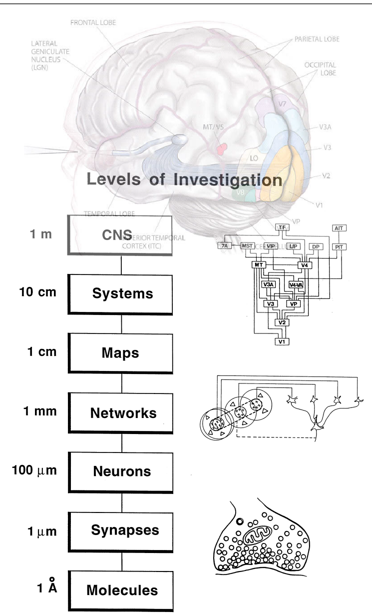
complex information processing system

where to start?

anatomy, neurophysiology… at what scale?

neurocomputational theory?

or behavior?



Visual behavior—jobs of vision

Within-object relations: Object perception

- categorization, identification
- 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,…

a ‘simple’ illustration

It takes just one quick glance to see the fox, a tree trunk, some grass and background twigs.

But that is just the beginning of what vision enables us to do with this picture.

Here’s one person’s description:

“One can see that there is an animal, a fox--in fact a baby fox. It is emerging from behind the base of a tree not too far from the viewer, is heading right, high-stepping through short grass, and probably moving rather quickly. Its body fur is fluffy, reddish-brown, relatively light in color, but with some variation. It has darker colored front legs and a dark patch above the mouth. Most of the body hairs flow from front to back...and what a cute smile, like a dolphin.”

Inferences about the fox picture involved various:

- 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 crucial assumption is that these inferences are based on deep, generative knowledge of how virtually any natural image could be produced

…after all, this may be the first time you’ve seen this picture!

how should one go about understanding perception?

computational problems?

Need to 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

computational problems?

Need to solve 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.

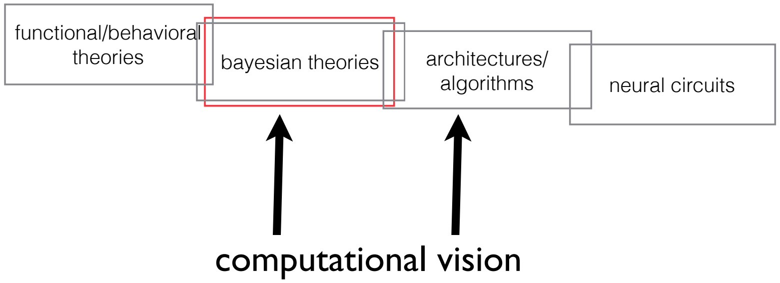
computational problems?

Need to solve task flexibility

Vision stimulates and support answers to a limitless range of questions. Human vision doesn't just recognize, it interprets scenes.

e.g. description of the fox

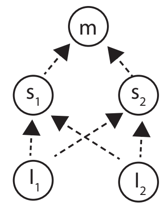
starting point for modeling?



bayesian decision theory provides framework for modeling uncertainty

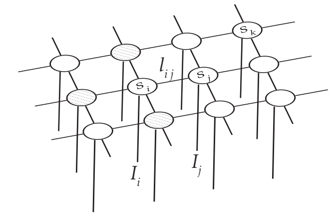
architectures/algorithms provide tools for understanding scalability and task flexibility

graphical models



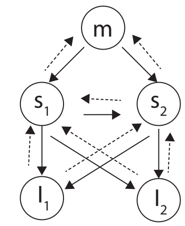
hierarchical organization

feedforward processing



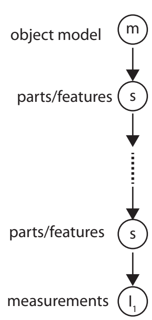
lateral organization,

lateral processing, reciprocal interactions between feature of similar type

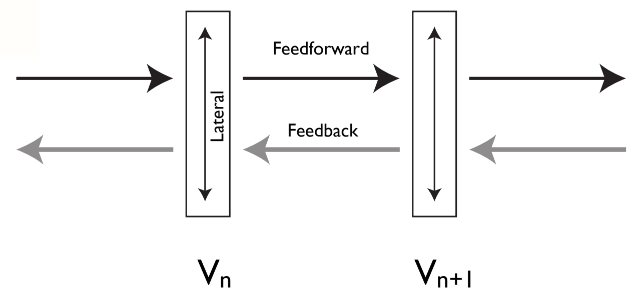


hierachical organization

feedforward, feedback & lateral processing



basis for hierarchies of abstraction

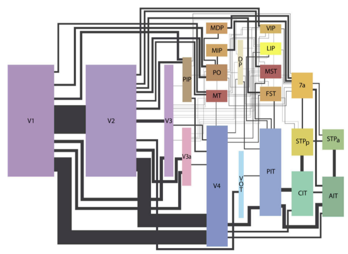


Are there common principles of organization and computation laterally, feedforward, and feedback?

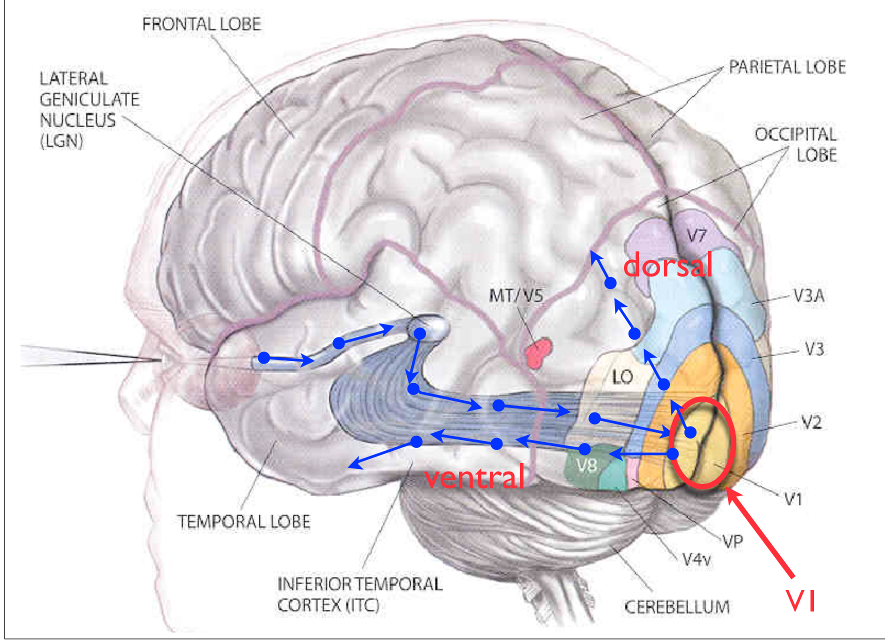
theories of the brain’s internal processes of perceptual inference

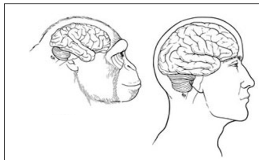
30+ cortical areas that are visually sensitive, often with specific preferences, such as

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

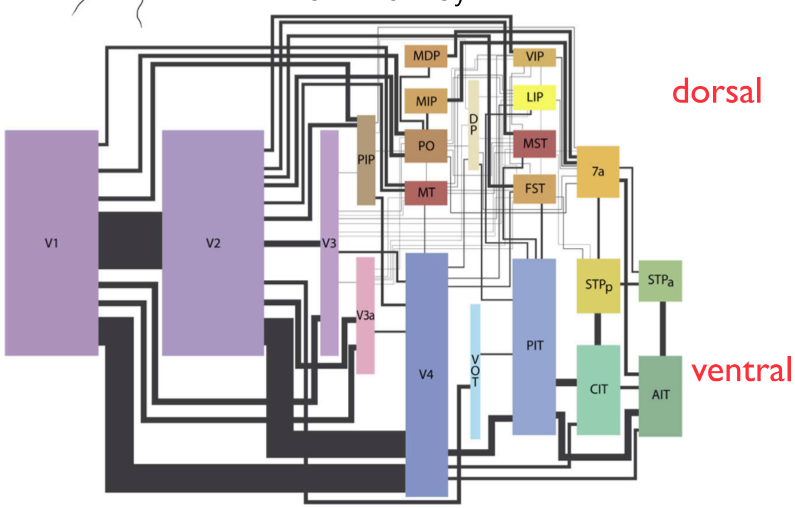


Wallisch, P., & Movshon, J. A. (2008). Structure and Function Come Unglued in the Visual Cortex. 197.





from monkey



primary visual cortex (V1)

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

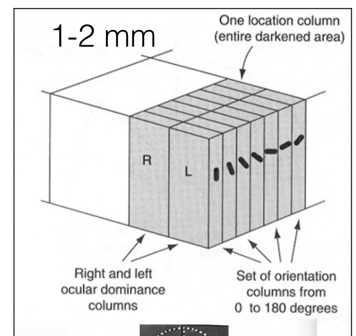
neurons representing similar features are near on cortical surface

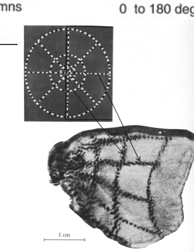
“simple cells” — template matching

“complex cells” — template matching tolerant to spatial shifts

global: hypercolumns arranged retinotopically

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

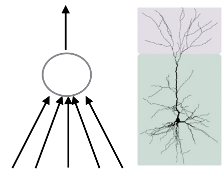




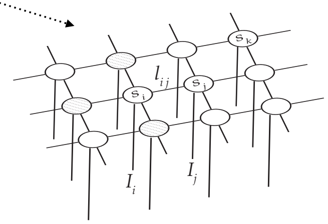
lateral organization: "maps"

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



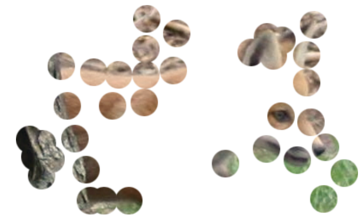
limited dendritic spread



Markov Random Field models

Grouping

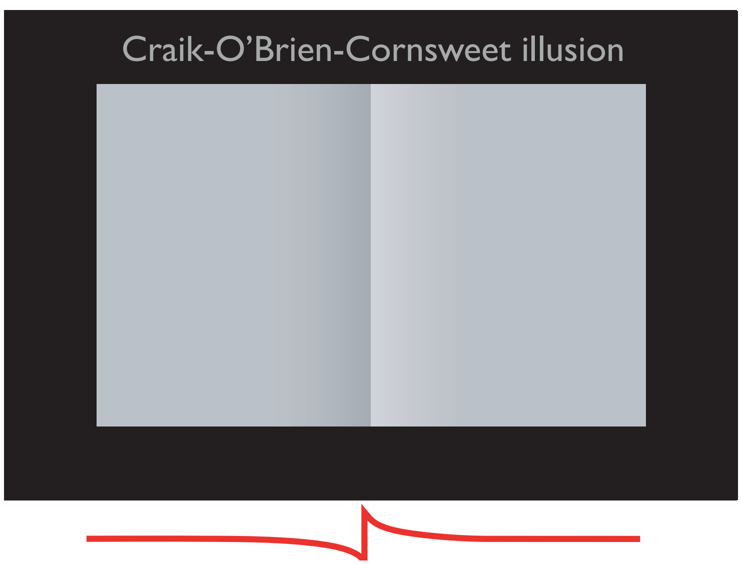
link contours with similar orientations

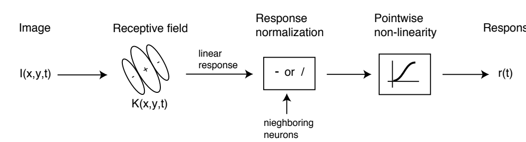


link regions with similar colors, textures

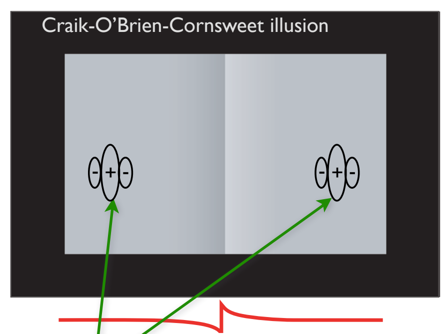
What should the local features be? How many different types?

An example





Convolution



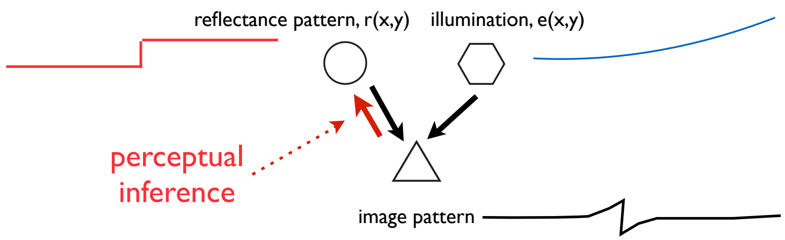
V1 response follows perceived lightness, not physical intensity

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

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

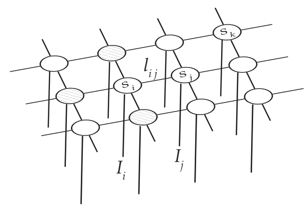
What are the features that are being linked?

image = f(pigment, illumination) ~ r(x,y) × 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



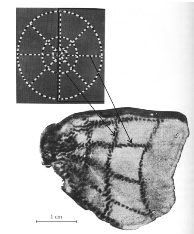
lateral organization

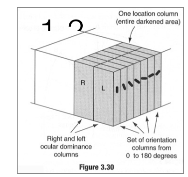
Why the organization? The level of abstraction?

- Keep similar features together for feedforward integration.
- Lateral computations to group features of similar type—segmentation
- Efficiency constraints
  - Minimum wiring constraint

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

  - Efficient representation of sensory input & cost of neural activity
  - Efficient representations for learning





how can layout be learned?

Kohonen adaptive maps

- Mathematica notebook

lateral organization

Why the organization? The level of abstraction?

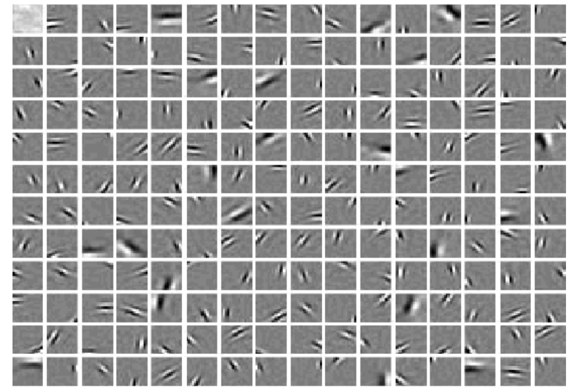
- Keep similar features together for feedforward integration.
- Lateral computations to group features of similar type—segmentation
- Efficiency constraints
  - Minimum wiring constraint
  - Efficient representation of sensory input & cost of neural activity
  - Efficient representations for learning

how can receptive field weights be learned?

both unsupervised, and supervised learning methods

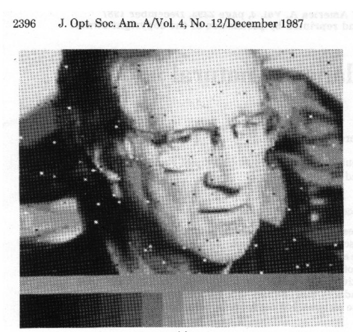
Unsupervised learning of receptive fields

- Models of the early levels of abstraction:
  - local, selectivity to orientation, spatial and temporal frequency
- Information-theoretic constraints
  - exploit regularities in natural image input



Olshausen & Field's model of V1 receptive fields

Efficient coding and higher order dependencies



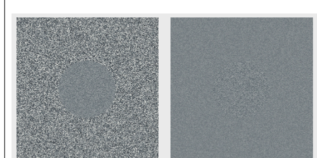
Kersten, D. (1987). Predictability and redundancy of natural images. J Opt Soc Am A, 4(12), 2395-2400.

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 $\sim L(x-1)$

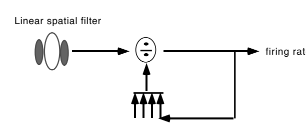
Divisive normalization--a common non-linearity



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.



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

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

Lateral organization

How do neural populations represent information?