Psy 5036 Lecture 25
Recognition in background clutter:

The role of top-down processing in segmentation & recognition

Object recognition, given real images

- clutter, occlusion, noise
- role of cortical architecture

Object recognition in real images

Background clutter and occlusion

Challenge of complexity in natural image input

- Enormous range of variability in the images for a given object category, eg. "foxes"
- Lots of types of objects
- Enormous objective uncertainty regarding local image features present for any given exemplar

Background/context can be useful

- Background can provide prior information. E.g."index" cues, to narrow down the space of possible objects to be recognized. Depth context can be important. Oliva et al. (2003), Torralba et al. (2006)
- For demonstrations of the role of background semantic content for human recognition:

  Biederman I (1972) Perceiving real-world scenes. Science 177:77-80.

Image patches corresponding to the car and person are the same except for 90 deg rotation

http://web.mit.edu/torralba/www/carsAndFacesInContext.html

Background & clutter can also be a problem

1. Segmentation is difficult because clutter near a target object's borders produce misleading edges.
2. There are also missing edges due to noise, lighting variation, and occlusion where other surfaces may cover parts of the target object

Object recognition given occlusion, clutter

Linking local information (features) likely to belong to the same object or pattern

- local ambiguity, noise
- need for good features & integration, generic priors, e.g. smoothness, contour and region-based grouping

Resolving competing explanations

- occlusion, clutter
- need for domain-specific priors

Good features, like color & stereo can help with segmentation ...but not always there, and we still need to make decisions about object categories

Strategies

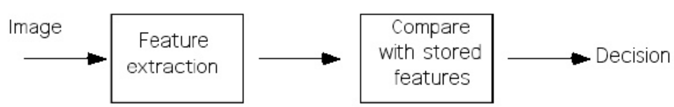
Discriminative mechanisms

- Use reliable low-level features. Computational/behavioral speed and accuracy requires effective diagnostic features to deal with the enormous with-class variation within a pattern/ object category

Generative mechanisms

- Provide flexibility

Discriminative models

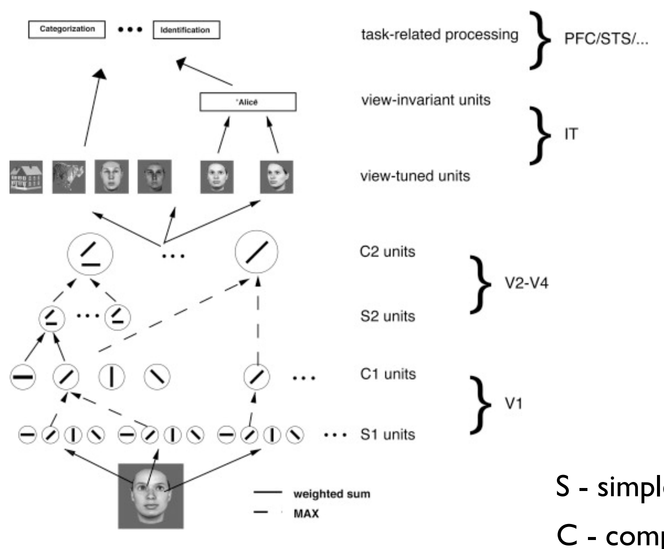


Bottom-up

Need to use/learn features to support reliable if not perfect first, bottom-up pass

Hierarchical models for feature extraction

- Local features progressively grouped into more structured representations
  - edges => contours/fragments => parts => objects
- Increased selectivity for object/pattern type
- Decreased sensitivity to view-dependent variations of translation, scale and illumination



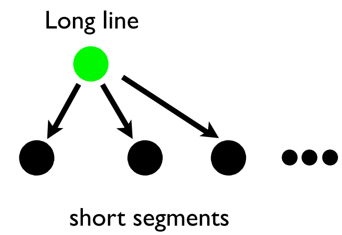
S - simple cell like

C - complex cell like

Riesenhuber & Poggio, 1999

Edges to contours

Cue integration

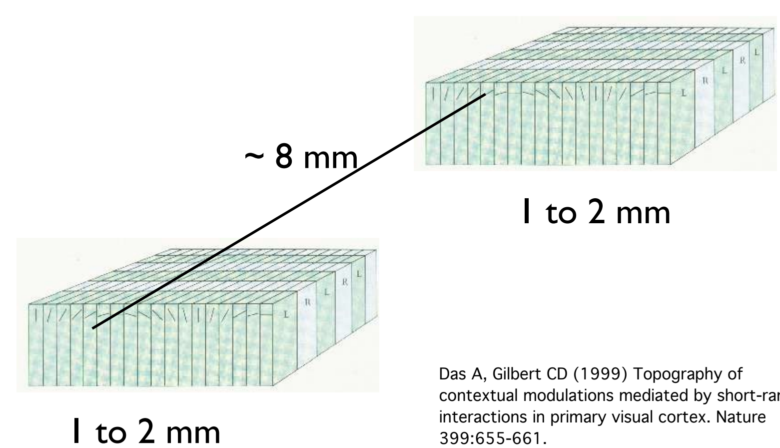


Parent P, Zucker SW (1989) Trace inference, curvature consistency, and curve detection. IEEE Transactions on Pattern Analysis & Machine Intelligence 11:823-839.

Yuille AL, Fang F, Schrater P, Kersten D (2004) Human and Ideal Observers for Detecting Image Curves. In: Advances in Neural Information Processing Systems 16 (Thrun S, Saul L, Schoelkopf B, eds). Cambridge, MA: MIT Press.

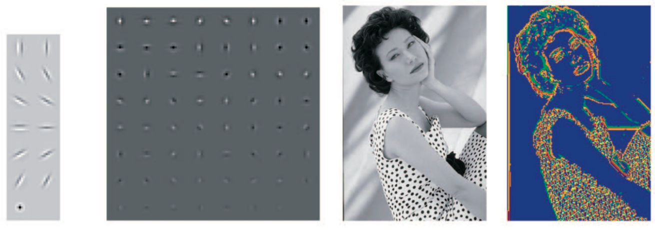
Cortical basis?

Short segments to long lines? Within-area linkage?



Das A, Gilbert CD (1999) Topography of contextual modulations mediated by short-range interactions in primary visual cortex. Nature 399:655-661.

What if the edges supporting boundaries are ambiguous? Region/texture-based grouping

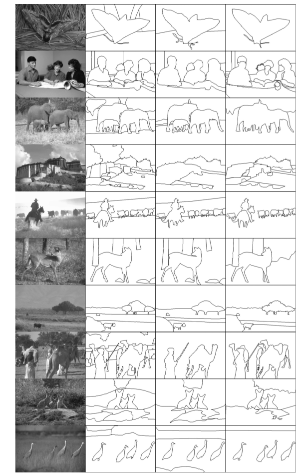


(a) (b) (c) (d)

From: Martin, D. R., Fowlkes, C. C., & Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Trans Pattern Anal Mach Intell, 26(5), 530-549.

Texture-based grouping

- "A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics" D. Martin, C. Fowlkes, D. Tal, and J. Malik. ICCV 2001
- “super-pixels”



Object recognition given occlusion, clutter

Linking local information (features) likely to belong to the same object or pattern

- local ambiguity, noise
- need for generic priors, e.g. smoothness of contours, color, texture statistics

Resolving competing explanations

- occlusion, clutter
- need for domain-specific features

Domain-specific features

- How to learn features to support a variety of actions, not just decisions about labels

How to learn features to support a variety of actions, not just decisions about labels?

- Size perception, e.g. for interception
- Material, e.g. for driving
- ...
- Object categorization
  - Do discriminative features learned in one task transfer to another?

Computational example: Learning informative features for a task

What do these scenes have in common?

With Evgeniy Bart

“Up” curbs-- that require a step up

Distinguish from Non- “up curbs”

But may require a different action

...that do not require a step

Selecting diagnostic features

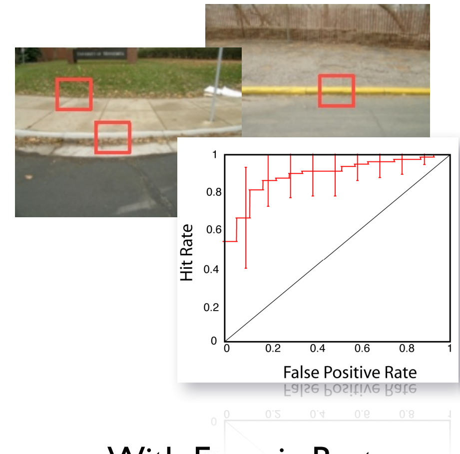
$$I(C;F) = H(C) - H(C|F)$$

$$\begin{aligned} F_1 &= \arg\max_F I(C;F); \\ F_{k+1} &= \arg\max_F \min_i I(C;F|F_i) \end{aligned}$$

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. Nat Neurosci, 5(7), 682-687.

Learning based on informative fragments for the task

- Find fragments that maximize mutual information (Ullman et al., 2002; Bart et al, 2004)
- Detect "up curbs" from an approach angle that requires a step



With Evgeniy Bart

Learning object categories

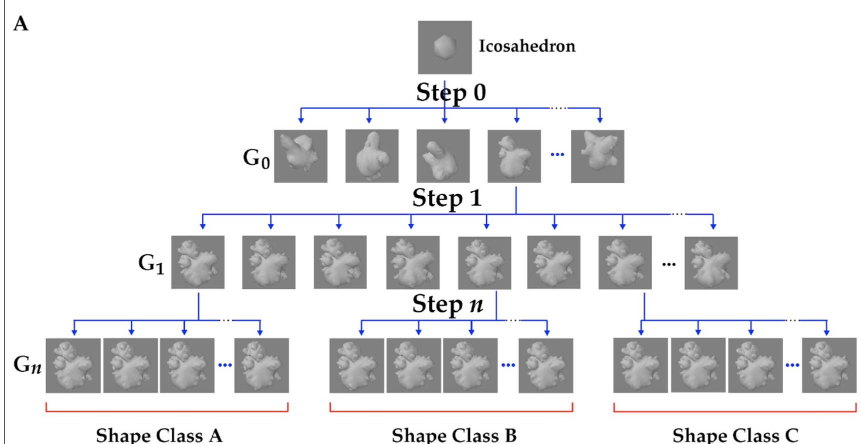
Do image features (fragments) that maximize mutual information predict the features that human observers learn to use?

Use novel object classes with small within-class variation and slightly larger between-class variation

Virtual phylogenesis of digital embryos

Hegde, J., Bart, E., & Kersten, D. (2008). Fragment-Based Learning of Visual Object Categories. Curr Biol. 18, 597-601

Virtual Phylogenesis



Training

A | A or B? | B

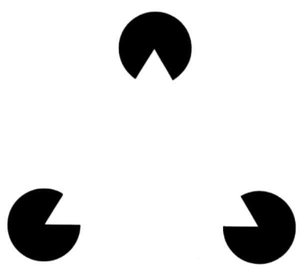
Results: Transfer of skill?

- For new previously unseen exemplars?
  - Yes. Human observer classification was predicted by features chosen to maximize mutual information. In other words, a set of features that were shared within a class, but at the same time most effective at discriminating classes

A

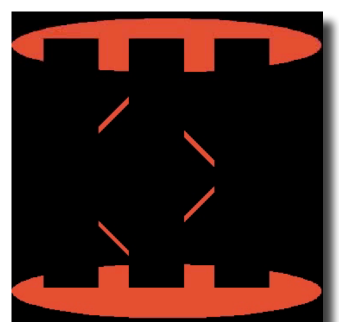
Hegde, J., Bart, E., & Kersten, D. (2008). Fragment-Based Learning of Visual Object Categories. Curr Biol. 18, 597-601

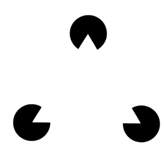
But what if ambiguity is high: missing edges and/or no texture?

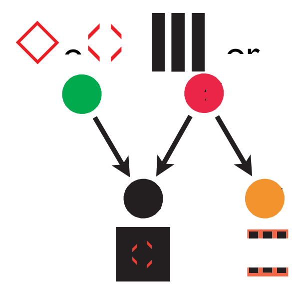


Top-down, generative models?

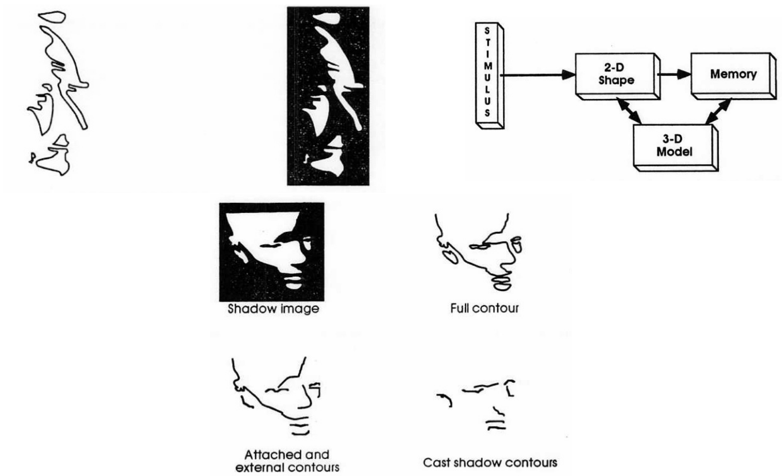
Solving "perceptual puzzles": Auxiliary evidence for occlusion





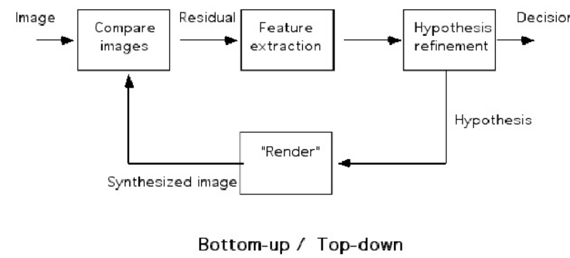


Recognition despite cast shadows

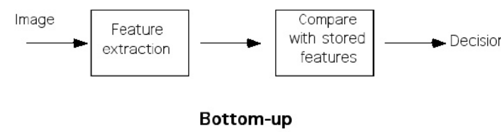


Cavanagh P (1991) What's up in top-down processing? In: Representations of Vision: Trends and tacit assumptions in vision research (Gorea A, ed), pp 295-304. Cambridge, UK: Cambridge University Press.

Uses of a generative model



In contrast with strictly feedforward



A generative model could be used in many ways. Generating prediction errors is just one. One could also use it to highlight lower-level features that are consistent with the high-level explanation.

- Doesn't mean that feedback is necessary for recognition
- Rather, top-down feedback used as needed
  - for achieving high-performance given uncertainty, noise, clutter
  - learning new object models

Computer vision
Image parsing: analysis by synthesis
(Tu, Z., Chen, X., Yuille, A., & Zhu, S. (2005))

- Find most probable scene description
- Bottom-up "proposals" (cues) to access three types of models (text, faces, background/texture) models
- Verification through top-down synthesis
- If bottom-up proposals are good, synthesis is not needed to find most probable scene
- Flexible graph

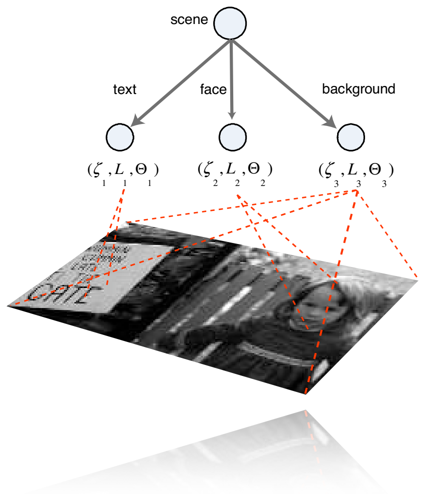
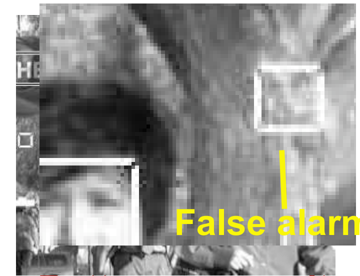


Image parsing & "Explaining away"

Input

Tu, Z., Chen, X., Yuille, A., & Zhu, S. (2005). Image Parsing: Unifying Segmentation, Detection and Recognition. IJCV, 63(2).

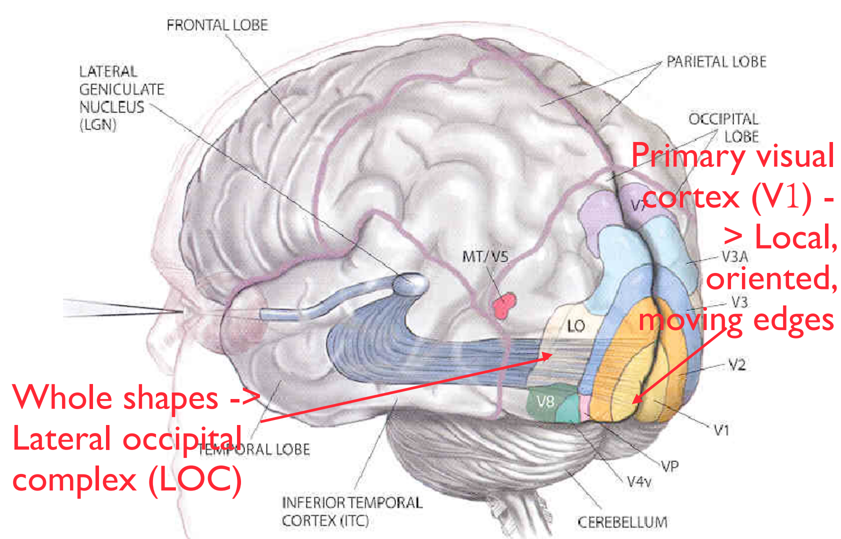


Bottom-up result

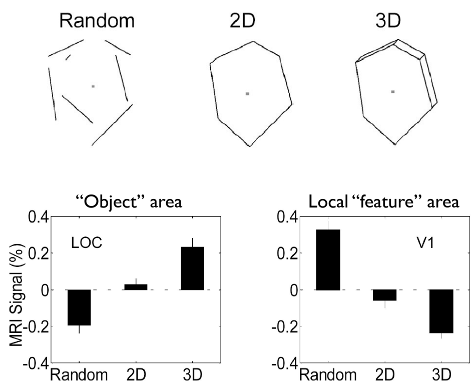
Next
Learning object categories
Spatial layout

Synthesized image

Neural evidence for top-down role in resolving ambiguity?



Shape perception can reduce V1 activity



Explanation?

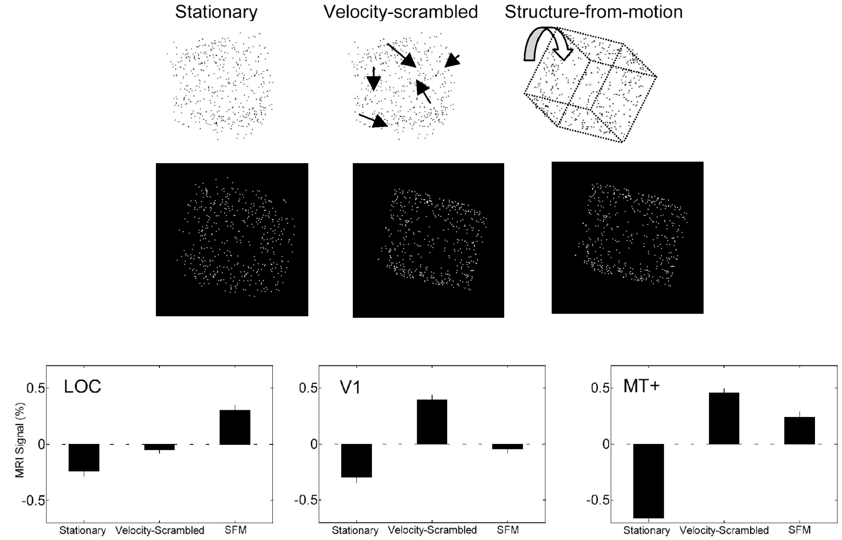
Many...

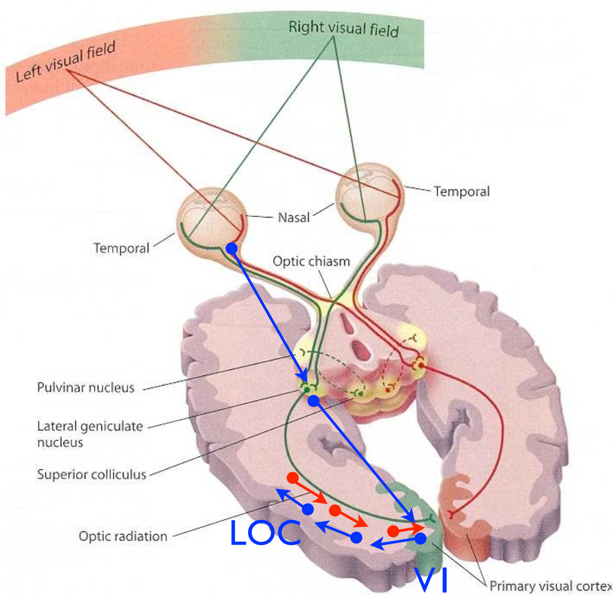
"Explaining away" through predictive coding

Sparse coding

Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. Proc Natl Acad Sci U S A, 99, 15164-15169.

Structure from motion





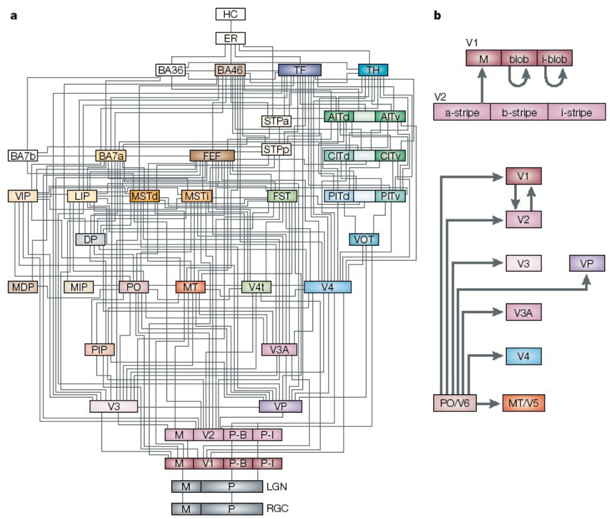
Cortical Mechanism?
...some speculation

1. Feedforward: local features to objects
2. Feedback models
   a. Feedforward + attention:
      competitive selection of features
   b. Predictive coding
   c. Sparsification

Internal generative models

MacKay DM (1956) The epistemological problem for automata. In: Automata Studies (Shannon CE, McCarthy J, eds), pp 235-250. Princeton: Princeton University Press.

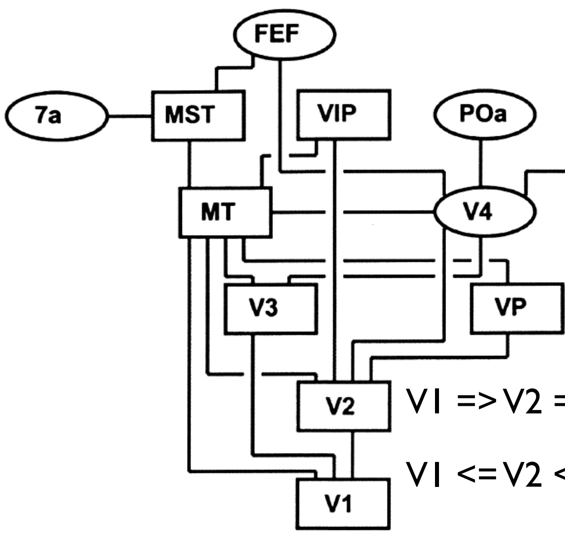
Cortical organization



Cortical organization

- Organization of visual cortices is a hierarchy
- Depends on distinct feedforward/feedback pathways
- Different laminar specificity
- More backward connections
- Backward connections more diffuse

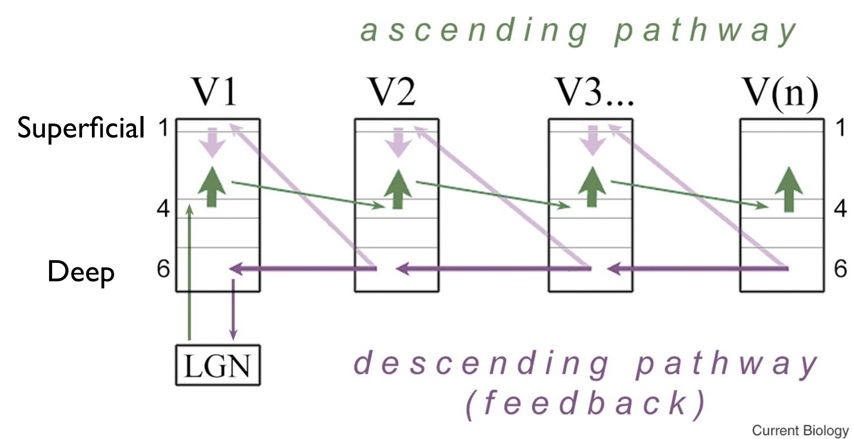
Object recognition?



V1 => V2 =>V4 =>IT

V1 <= V2 <=V4 <=IT

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

Forward connections

- Sparse axonal bifurcations
- Topographically organized
- Originate in supragranular layers (I,II,III)
  - III => adjacent columns
  - II => other cortical areas
- Terminate in layer IV

Friston K (2003) Learning and inference in the brain. Neural Netw 16:1325-1352.

Feedback connections

- Lots of axonal bifurcation
- Diffuse topography
- Originate in infragranular (V,VI) layers
- Mainly terminate in supragranular layers (I,II,III)

Friston K (2003) Learning and inference in the brain. Neural Netw 16:1325-1352.

Markov, N. T., & Kennedy, H. (2013). The importance of being hierarchical. *Current Opinion in Neurobiology*, 1–8. doi:10.1016/j.conb.2012.12.008

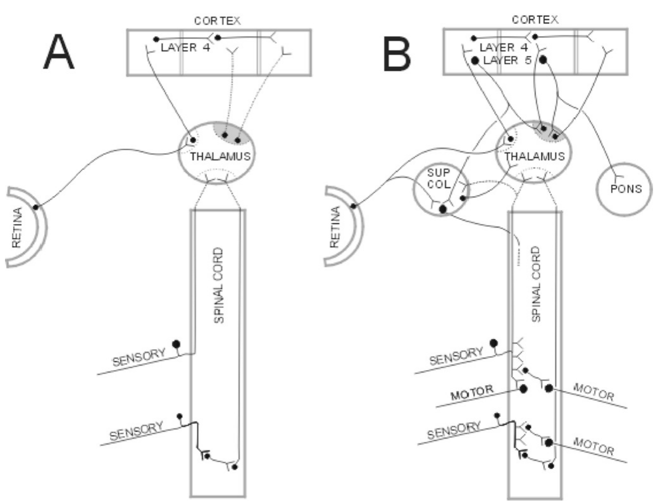


Figure courtesy of Ray Guillery

Internal generative models

Analysis-by-synthesis

Predictive coding

- High-level object models project back predictions of the incoming data

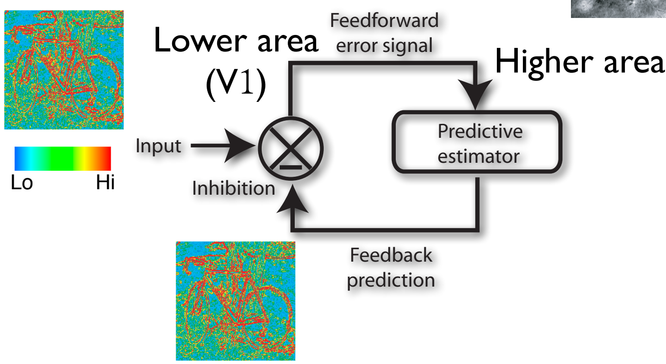
  Poor fit, high residual => high activity

Sparsification

- A good high-level fit tells earlier areas to “stop gossiping”

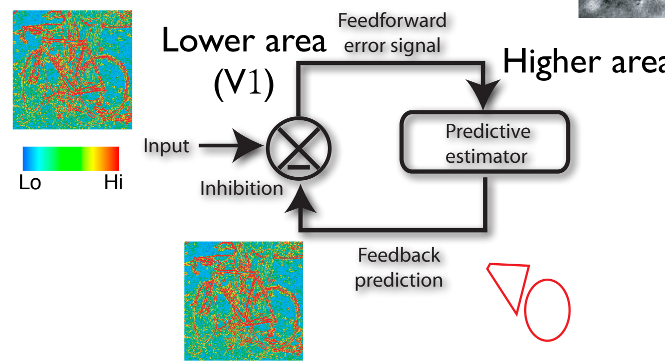
  Amplify the activity for early features that belong to object, suppress the rest

Predictive (top-down) processes in the brain?

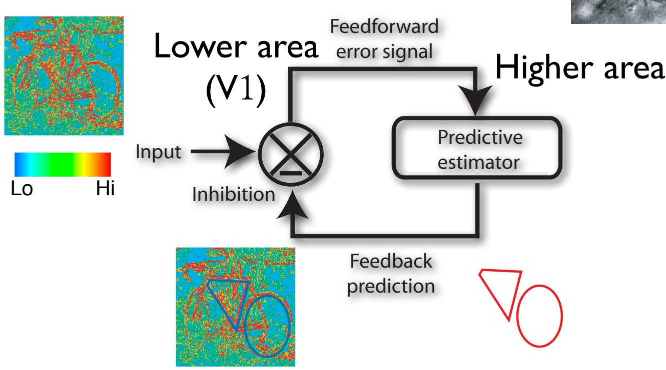


e.g. Rao, R. P., & Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. Neural Comput, 9(4), 721-763.

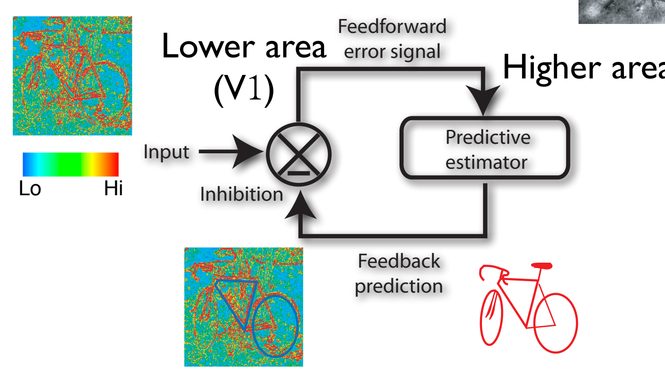
Predictive coding

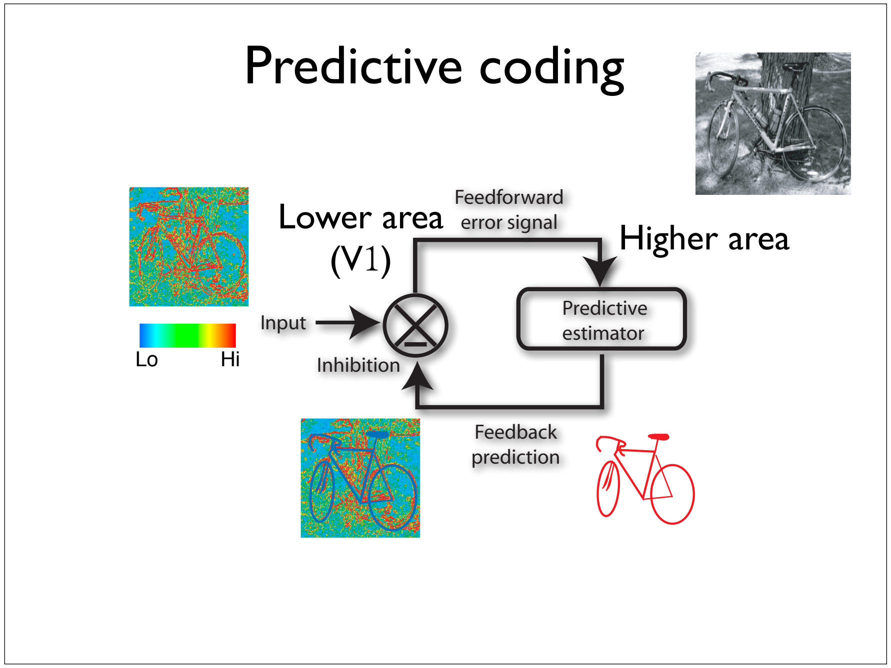


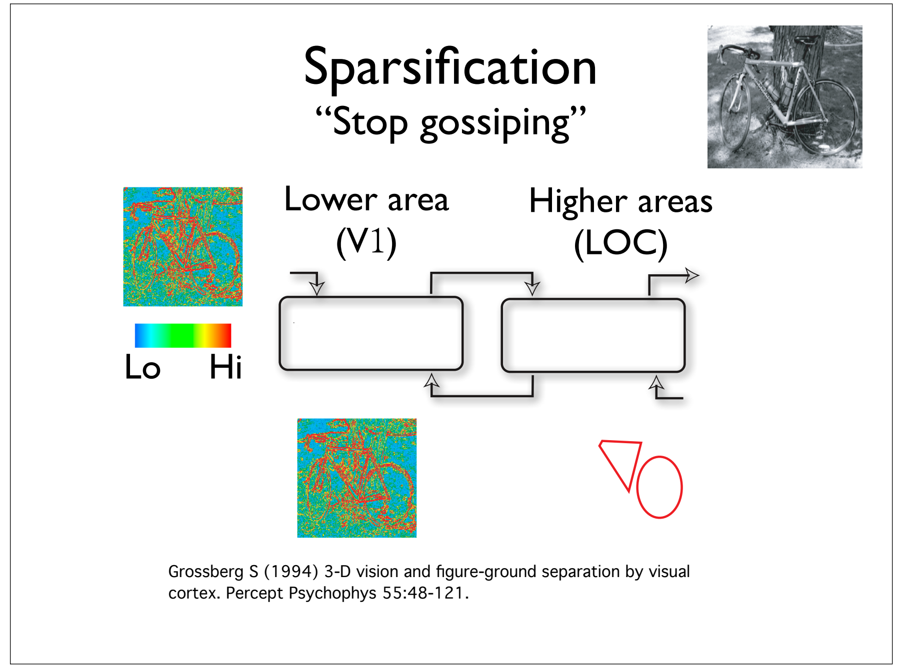
Predictive coding

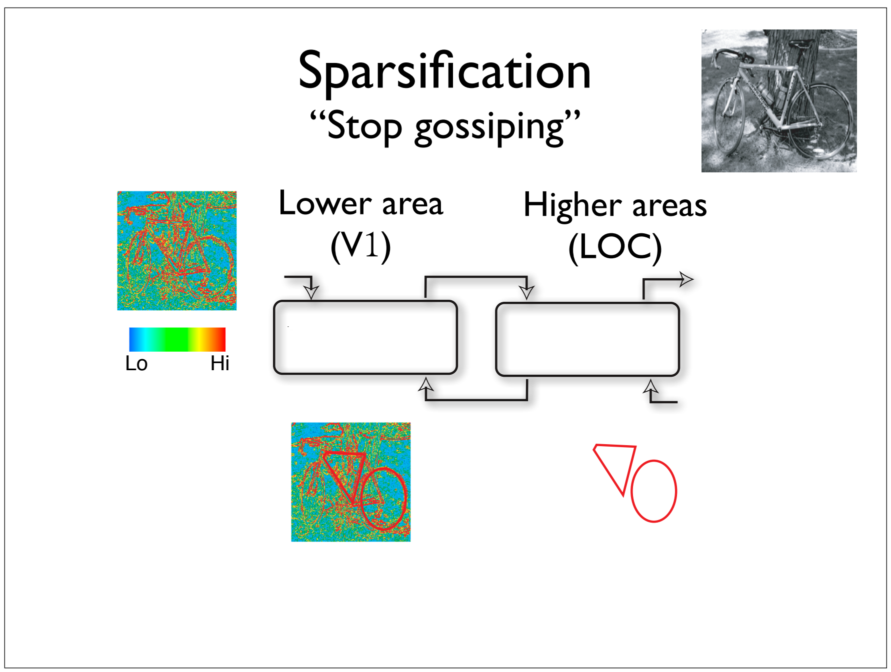


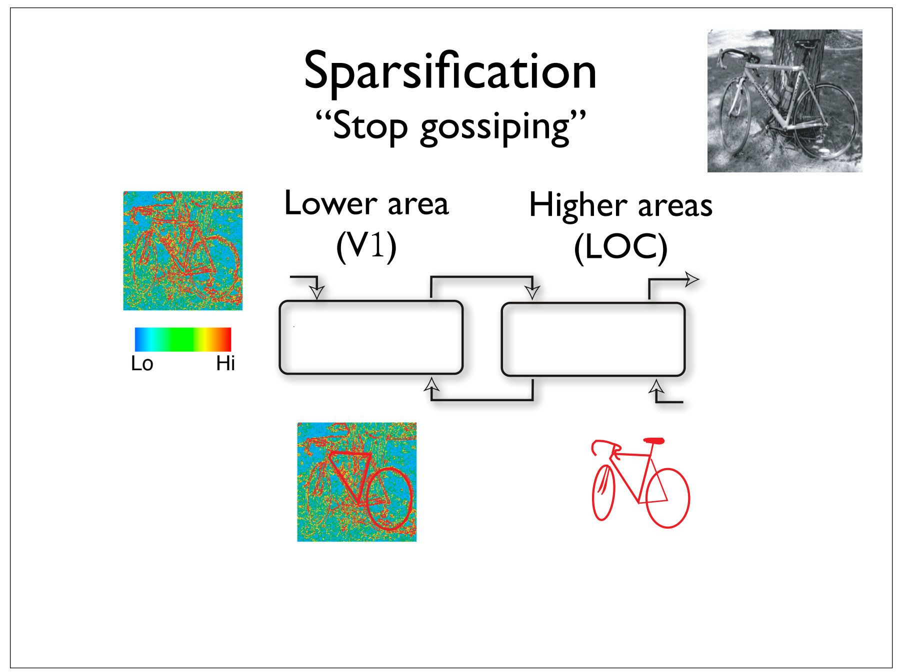
Predictive coding




Predictive coding
Lower area
(V1)
Feedforward
error signal
Higher area
Input
Predictive
estimator
Inhibition
Lo
Hi
Feedback
prediction

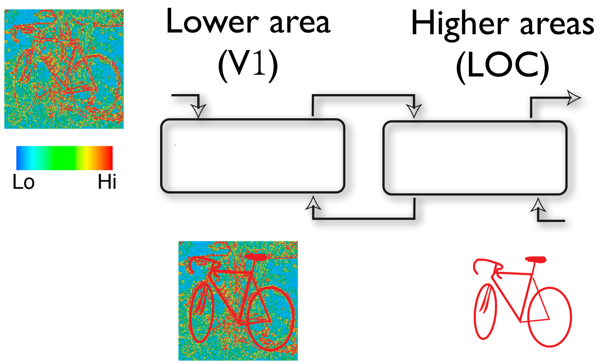

Sparsification
“Stop gossiping”
Lower area
(V1)
Higher areas
(LOC)
Lo
Hi
Grossberg S (1994) 3-D vision and figure-ground separation by visual cortex. Percept Psychophys 55:48-121.


Sparsification
“Stop gossiping”
Lower area
(V1)
Higher areas
(LOC)
Lo
Hi


Sparsification
“Stop gossiping”
Lower area
(V1)
Higher areas
(LOC)
Lo
Hi

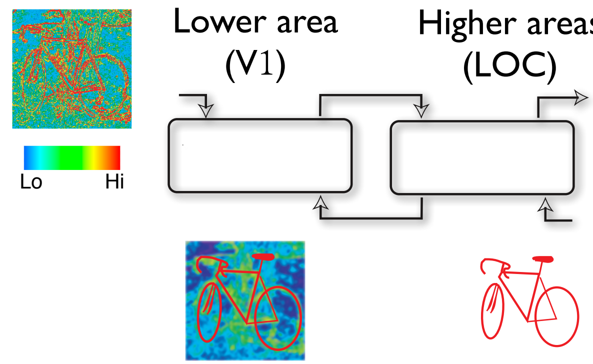
Sparsification

"Stop gossiping"

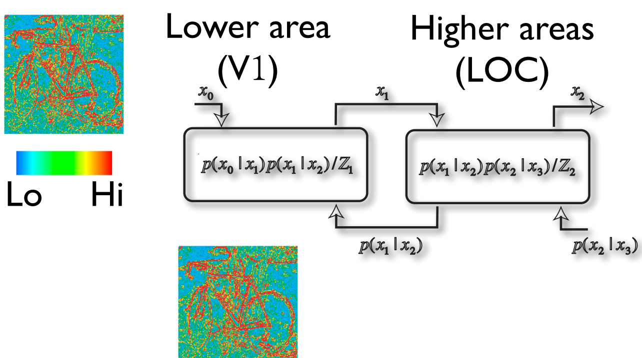


Sparsification

"Stop gossiping"



Bayesian Interpretation Sparsification



Lee & Mumford, 2003, JOSA

Particle filtering ideas: Isard M, Blake A (1998) Condensation -- conditional density propagation for visual tracking. International Journal of Computer Vision 29:5--28.

Zhu, L., & Yuille, A. L. (2011). A hierarchical compositional system for rapid object Zhu, L., Chen, Y., & Yuille, A. (2011). Recursive Compositional Models for Vision: Description and Review of Recent Work. *Journal of Mathematical Imaging and Vision*, *41*(1-2), 122–146. doi:10.1007/s10851-011-0282-2.

Summary

Common patterns of neocortex structure

- Has inspired lots of models of cortical information processing

Key target problem?

- Object perception/recognition given occlusion, clutter

fMRI and object grouping given occlusion

- consistent with feedback, but...