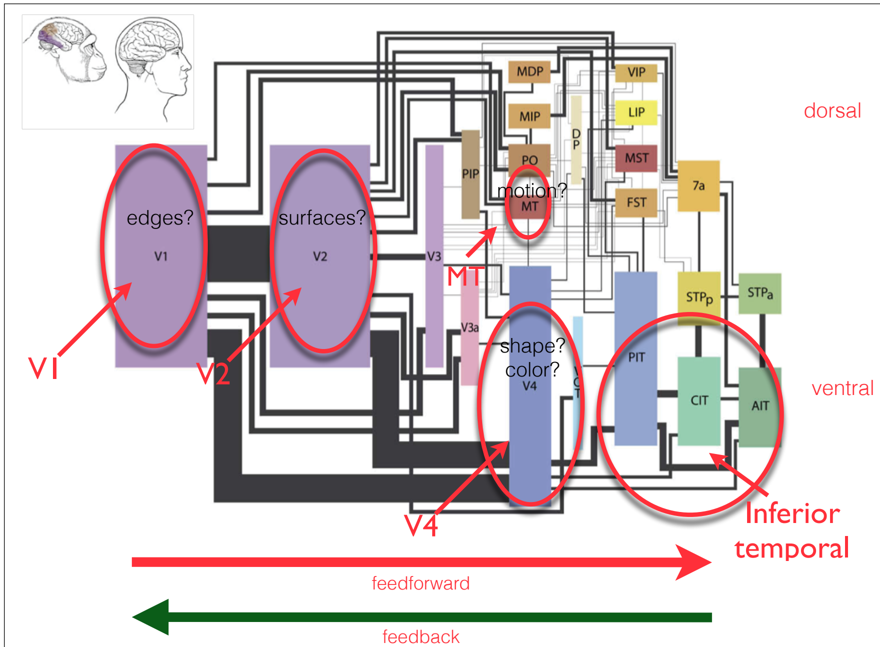
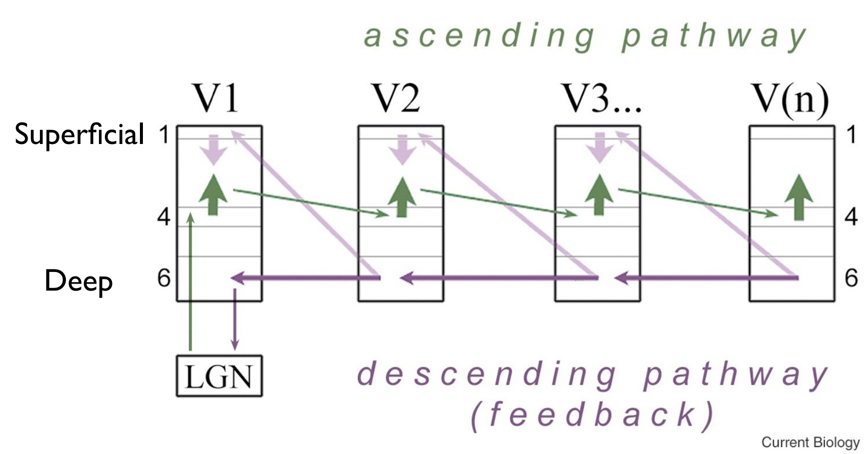
Bidirectional processing II:

feedforward & feedback networks for recognition

Focus on feedback computations

Feedback





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

Bayesian perspective: two computational strategies

Discriminative mechanisms — p(object | image) feedforward

- Computational/behavioral speed and accuracy requires effective diagnostic features to deal with the enormous variation within a pattern/object category

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*(4), 454–461.

Generative mechanisms — p(image | object) x p(object)* feedback

- Provide flexibility, generalization

* recall bayes: $p(\text{object} \mid \text{image}) \propto p(\text{image} \mid \text{object}) \times p(\text{object})$

Can feedback help with the local uncertainty, scalability and flexibility problems

Fine-scale recognition and segmentation

Unfamiliar objects/appearances

Learning given only a few examples

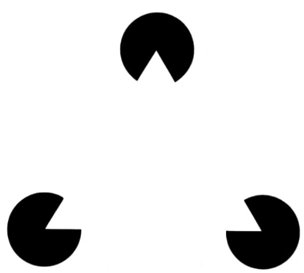
Bootstrap learning problem:
How to learn when objects aren't experienced in isolation?

Domain-specific compositional models

Automatic or consciously driven?

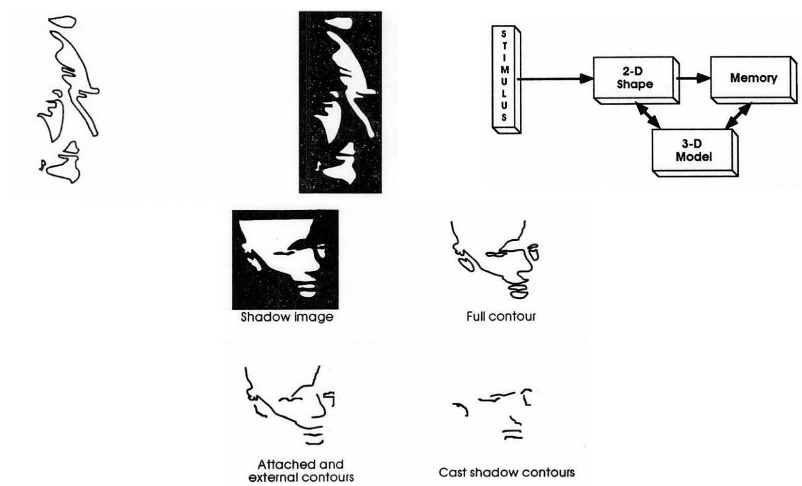
The executive metaphor
expertise at various levels of abstraction

local uncertainty, missing data

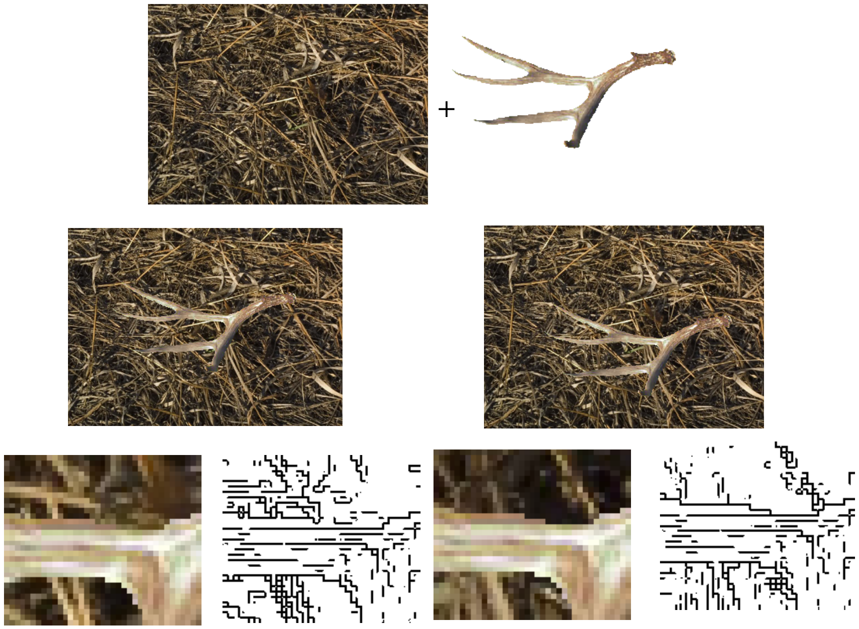


Top-down, generative models?

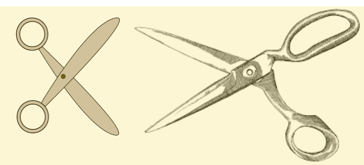
Extraneous data: recognition despite cast shadows



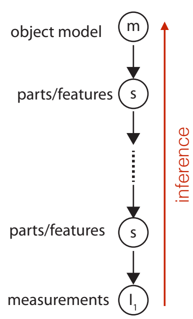
Cavanagh P (1991) What's up in top-down processing? In: Representations of Vision: Trends and tacit assumptions in vision research (Gorea A, ed), pp 295-304. Cambridge, UK: Cambridge University Press.



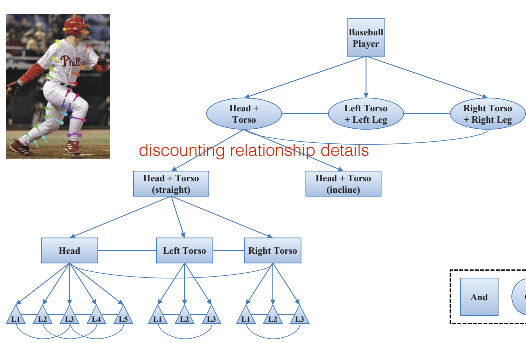
Object variations that haven't been seen before

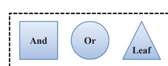


can recognize as scissors AND estimate an articulation



Object variations that haven't been seen before: Compositional architectures for representation





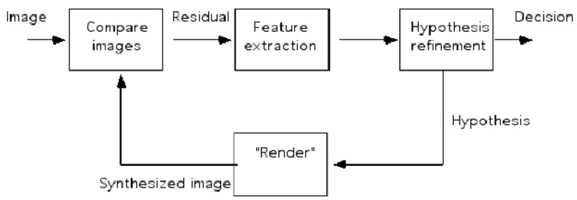
basic logical operation: detect "disjunctions of conjunctions"

Doesn't mean that feedback is necessary for recognition (Thorpe et al.)

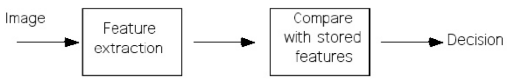
But top-down feedback may be important for

- achieving high-performance given uncertainty, noise, clutter
- task flexibility
- learning new object models

Contrast predictive coding with strictly feedforward



Bottom-up / Top-down



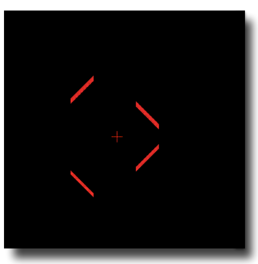
Bottom-up

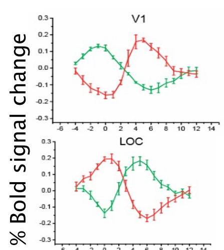
Disambiguation?

Predictive coding: suppress lower-level features that are consistent with a confident high-level interpretation. Reduce metabolic costs, signal new unexplained incoming information.

Analysis-by-synthesis. Bind lower-level information that might be required for executive tasks, e.g. fine-grain. : enhance lower-level consistent features and/or suppress inconsistent ones. Useful for representation and interpretation of novel patterns? Dealing with clutter?

perceptual organization reduces activity in V1



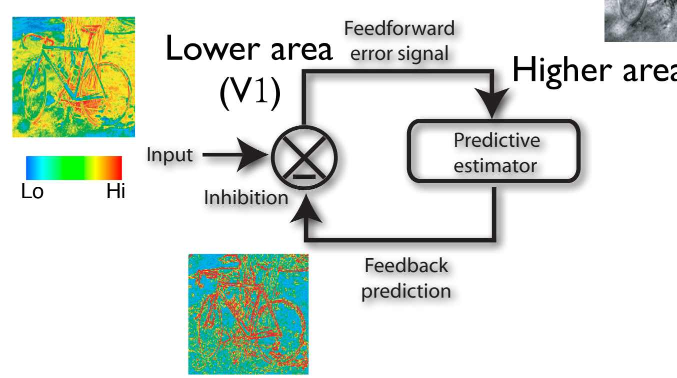


Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002).; Fang, F., Kersten, D., & Murray, S. O. (2008).

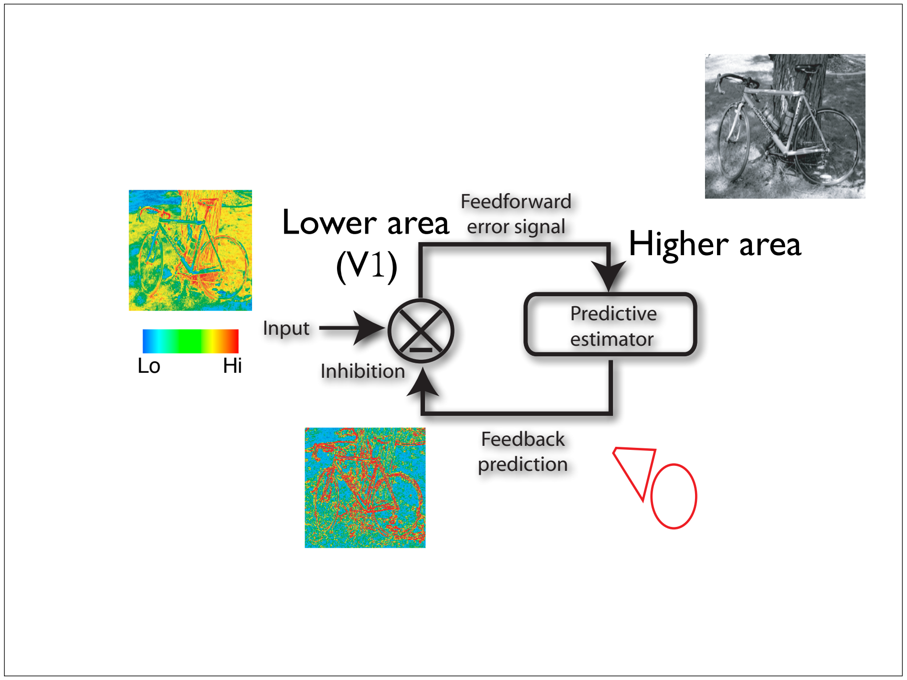
...but non-retinotopic voxels are also suppressed (Wit et al., 2012)

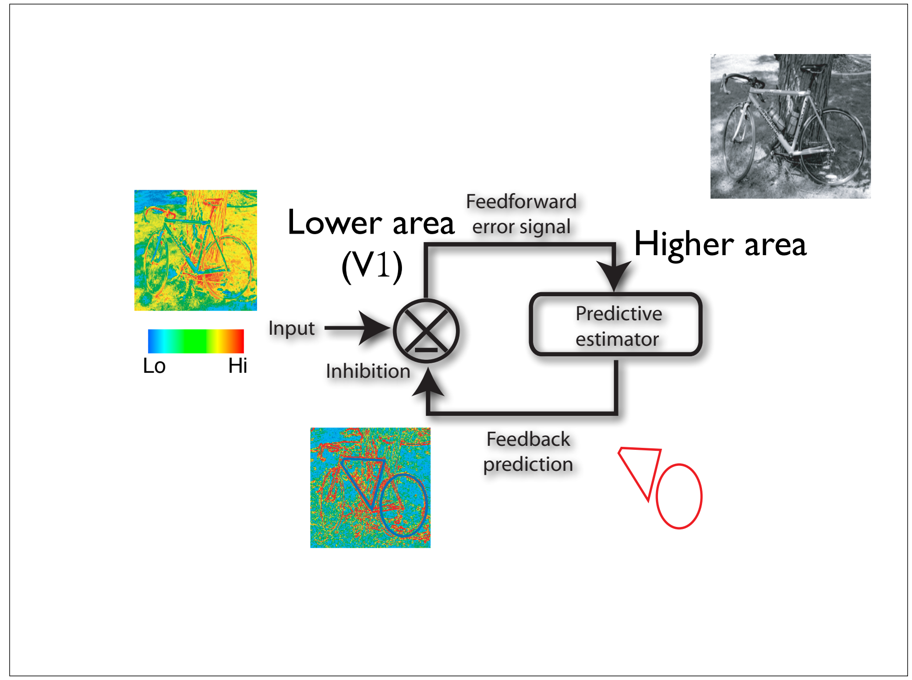
Behavioral evidence for top-down reduction of early activity? Use perceptual adaptation--the psychophysicist's electrode

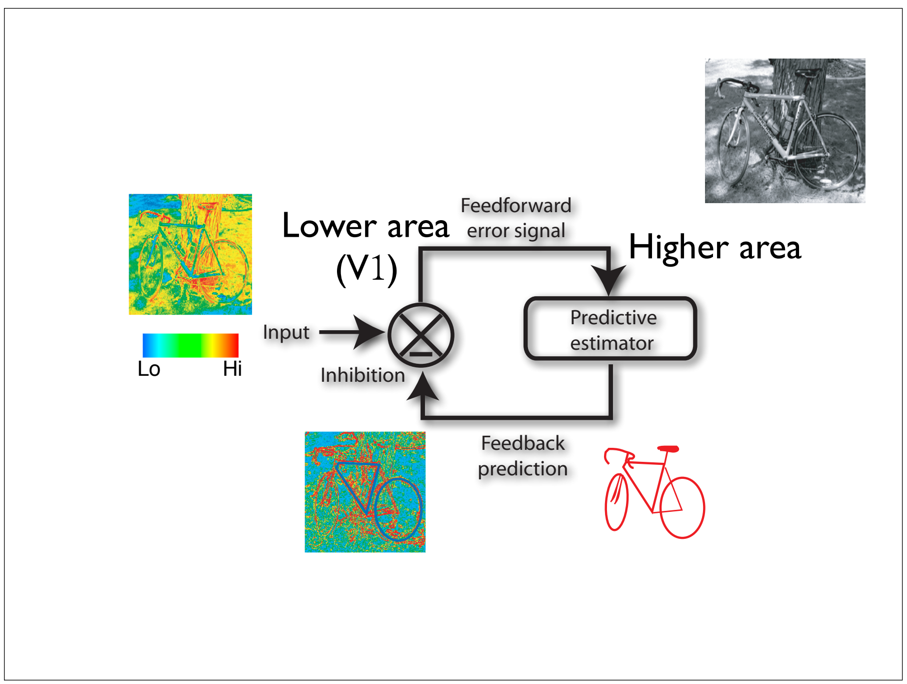
"predictive coding" through suppression of consistent features at lower levels

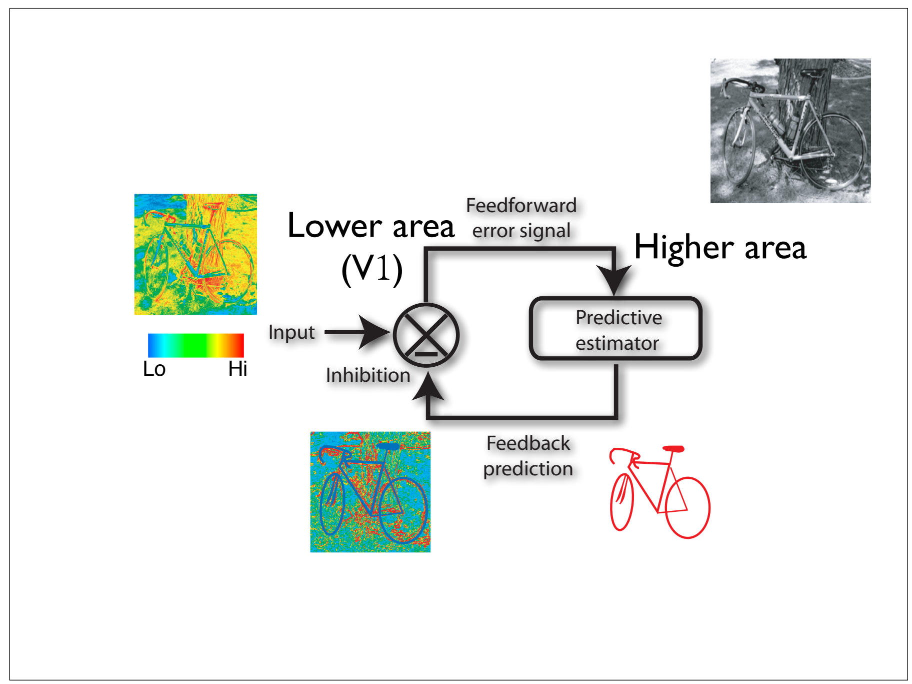


e.g. Rao, R. P., & Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. Neural Comput, 9(4), 721-763.

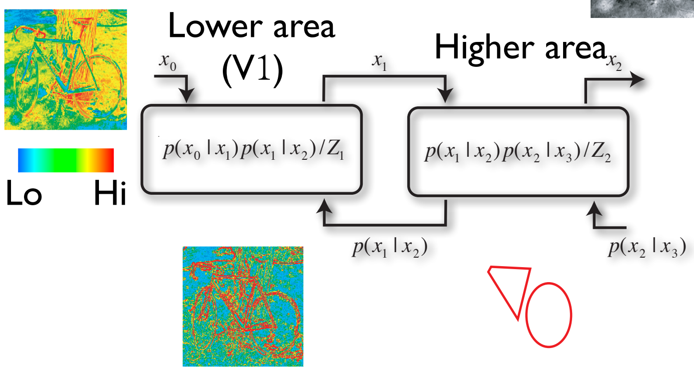

Lower area
(V1)
Feedforward
error signal
Higher area
Input
Predictive
estimator
Lo
Hi
Inhibition
Feedback
prediction


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

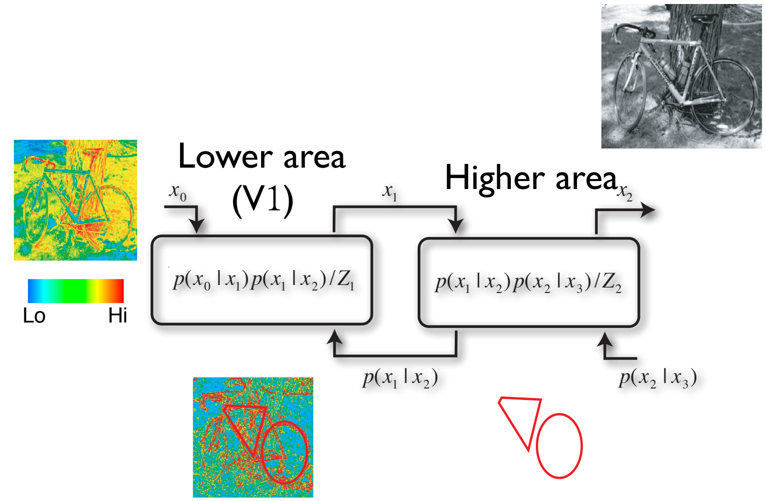

Lower area
(V1)
Feedforward
error signal
Higher area
Input
Predictive
estimator
Lo
Hi
Inhibition
Feedback
prediction

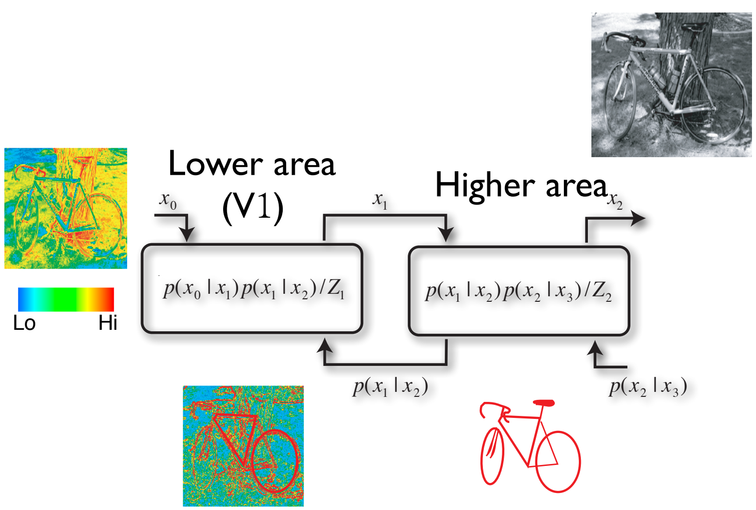

Lower area
(V1)
Feedforward
error signal
Higher area
Input
Predictive
estimator
Lo
Hi
Inhibition
Feedback
prediction

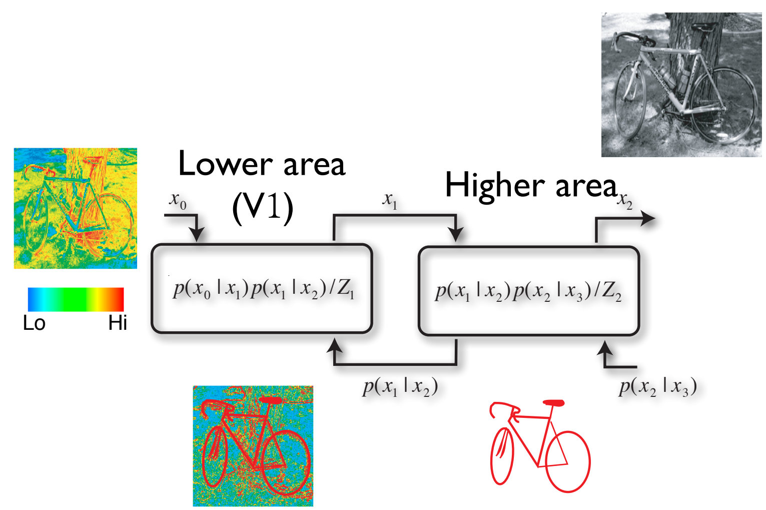
binding through enhancement
of consistent features at lower levels

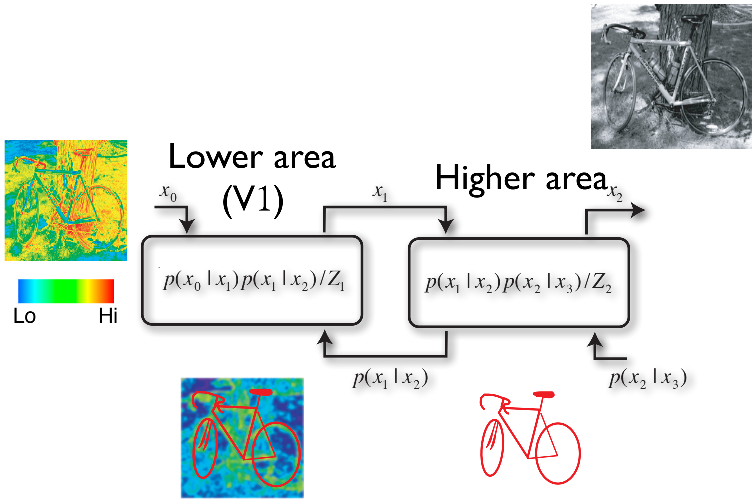


Lee & Mumford, 2003, JOSA





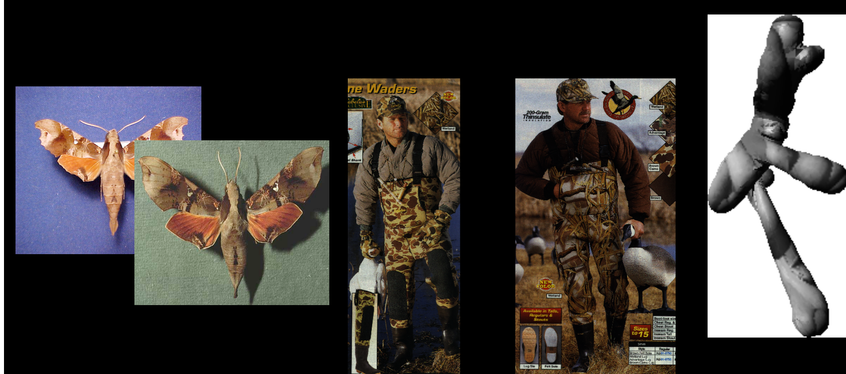




Return to the challenge of task flexibility

Humans can not only localize and recognize object categories, they can

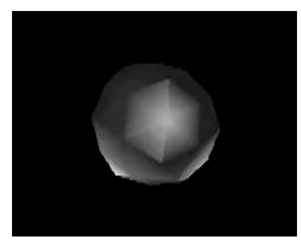
- parse, describe and precisely segment an image, and lots more, such as measure attributes and relations, infer intent, …
- rapidly learn new object models under difficulty segmentation conditions



Bringing nature to the lab. Callionima moths on the left show disruptive camouflage patterns. The garment in the middle image replicate images of the woodland background. Texture mapping was used by Brady and Kersten (2003) to mimic this for the digital embryo on the right. This introduces false positive object boundaries, and apparent shape from shading cues.

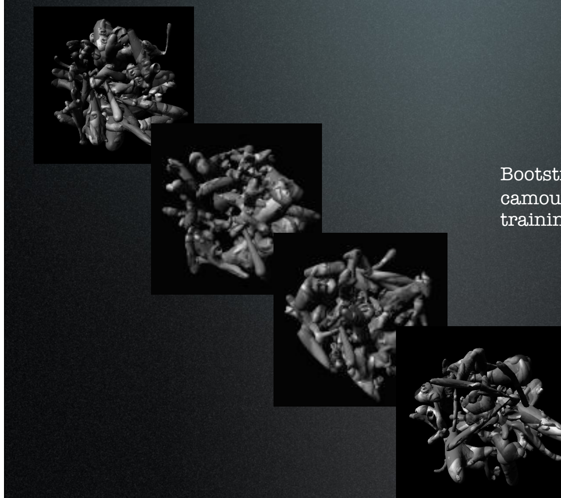
Virtual morphogenesis

Brady, M. J., & Kersten, D. (2003).
Bootstrapped learning of novel objects.
Journal of Vision, 3(6), 413–422.



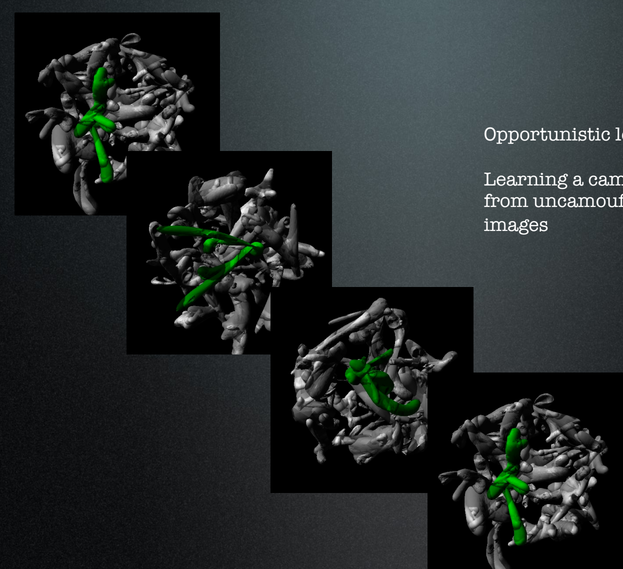
Brady, M. J., & Kersten, D. (2003).
Bootstrapped learning of novel objects.
Journal of Vision, 3(6), 413–422.

How do humans acquire prior knowledge of object classes? There is a target object in “plain view” in this figure. Without training, it is impossible to detect or draw a line around its boundary.



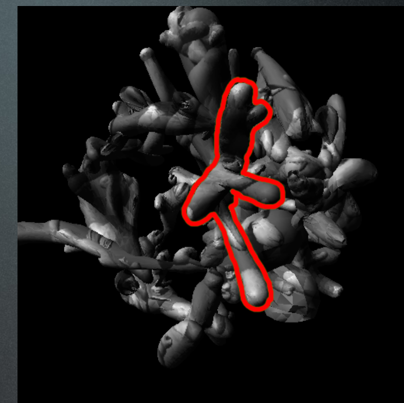
Bootstrapped learning: Learning a camouflaged object from camouflaged training images

If an observer has to opportunities to see colored birds, this could help the observer to learn about the forms that birds can take. Then at some future time, it could use this knowledge to see birds whose color does not distinguish it from the background, e.g. a different kind of bird, or under more difficult viewing conditions, such as during the night or fog. This is “opportunistic learning”



Opportunistic learning:

Learning a camouflaged object from uncamouflaged training images



After training, observers were tested on test images in which the objects were given new camouflage, and presented against new backgrounds.

All observers were able to learn opportunistically, and some were also able to learn from the camouflaged training images. This figure shows a perfect segmentation by an observer after training.

Flexibility

Limitations to current recognition algorithms as models of biological/human vision?

Humans generalize far beyond training data to novel images/ forms

To what extent does human visual flexibility, ability to generalize rely on deep generative knowledge?

How deep?

http://www.pauldebevec.com

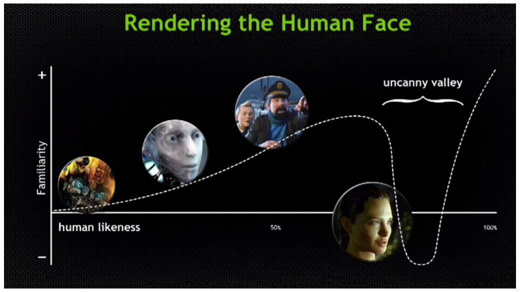
emotion/intent — muscles — multiple layers of soft tissue — 3D rendering parameters — Image

← deeper

Insights from computer graphics…

Take a look at faces, materials such as *hair and fluids*, and *body pose*

Message from computer graphics is as deep as you can given processing limitations



https://developer.nvidia.com/faceworks

General message for human visual neuroscience is “deep, but not too deep”.

“How to cheat and get away with it?”

Faces

See too: Nvidia talk facial expressions

How deep?

emotion/intent | muscles | multiple layers of soft tissue | 3D rendering parameters | image

deeper

Illumination?
Sources, shadows, inter-reflections,..

3D shape?

Material, e.g. sub-surface scattering?

2D shape approximations?

Appearance approximations?

image

deeper

Hair

hair care products have the highest sale volume of all non-food items in the US

What does it take to generate realistic hair?

Sasquatch software from www.worley.com

Hair can be…

- wavy, curly, straight, spiky, stiff, buzzed, shaved, parted, neatly-combed, tamed, long, short, cropped
- thick, full, lustrous, bushy, coarse, wiry
- thin, scraggly, fine, baby-fine, wispy, limp, flat, balding, receding
- black, brunette, brown, chestnut-brown, honey-blond, blond, golden-blond, ash-blond, auburn, red, strawberry-blond, gray, silver, white, salt-and-pepper
- permed, dyed, bleached, highlighted, weaved
- braids, ponytail, pigtails, bun, twist, bob, ringlets, flip, bangs, buzz
- layered, feathered, chopped, gelled, spiked, slicked down
- terminal and vellus

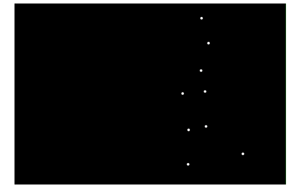
Viscous fluids

Body pose, actions

Toshev, A., & Szegedy, C. (2013). Deeppose: Human pose estimation via deep neural networks. arXiv Preprint arXiv:1312.4659.

X. Chen and A.L. Yuille. Articulated Pose Estimation with Image-Dependent Preference on Pairwise Relations. NIPS 2014

global



http://astro.temple.edu/~tshipley/ptlt_movies/mlwalk2.mov

http://www.biomotionlab.ca/Demos/BMLwalker.html

local

how to get from local to global?

Current inferential models of human visual recognition are not very "deep" in the sense of relying on inductive biases, generative models that could allow rapid learning from few samples, the ability to deal with almost any image (familiar or not).

Need to understand the critical dimensions that avoid the uncanny valley without computations and representations unlikely to exist in the brain. I.e. the "right" kind of generative model.

Need to understand how to model statistical regularities in classes of natural images. Linear methods are inadequate.

Need for compositional models, grammars, e.g. "recognition-by-components"