Bidirectional processing I:

feedforward & feedback networks for recognition

Focus today on feedforward architectures

recognition

- The computational problems of scalability and flexibility
- Feedforward models
- Feedback models

computational problems

Inferences about the image involve various inferences:

- types of features & attributes (shapes, material)
- recognition over 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

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.

In object recognition, humans have the capacity to *quickly* 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.

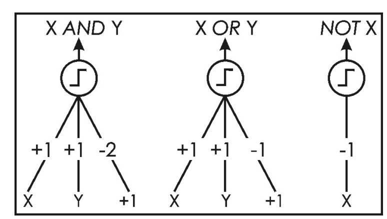
e.g. description of the fox

"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."

A little history of computational pattern/object recognition

1940s

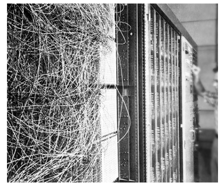
McCulloch and Pitts threshold logic units



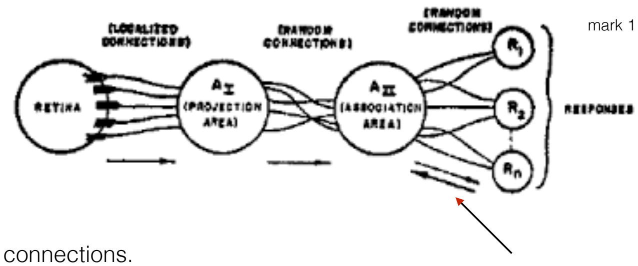
template models, e.g. SDT

1950s

Rosenblatt's perceptron



mark 1



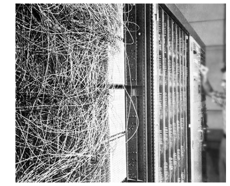
feedback too

random connections.
perceptron learning algorithm enabled it to
to learn to classify

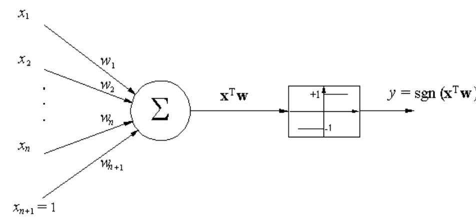
Rosenblatt, F. 'The Perceptron, a Perceiving and Recognizing Automaton', Cornell Aeronautical Laboratory Report No. 85-460-1 (1957);

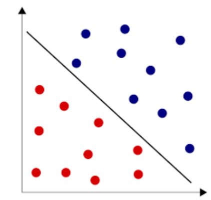
Rosenblatt, F. Principles of Neurodynamics (Washington, D.C.: Spartan, 1962).

Threshold-logic and the perceptron learning rule



mark 1



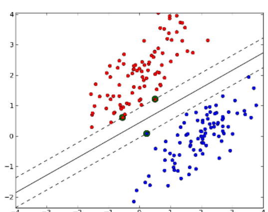


adjust weights, w, to find separating line
limited to linearly separable classification

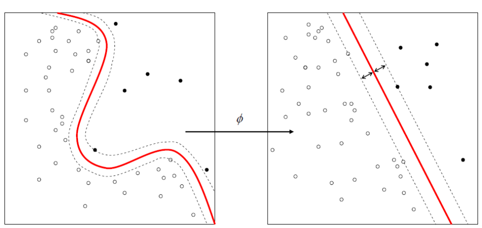
Rosenblatt, F. 'The Perceptron, a Perceiving and Recognizing Automaton', Cornell Aeronautical Laboratory Report No. 85-460-1 (1957);

Rosenblatt, F. Principles of Neurodynamics (Washington, D.C.: Spartan, 1962).

support vector machines



1963 — linear
1992 — non-linear kernels



$$\hat{y} = \mathrm{sgn} \sum_{i=1}^{n} w_i y_i k(\mathbf{x}_i, \mathbf{x}')$$

https://en.wikipedia.org/wiki/Kernel_method#/media/File:Kernel_Machine.png

1980s through 1990s
getting multi-layer perceptrons to work

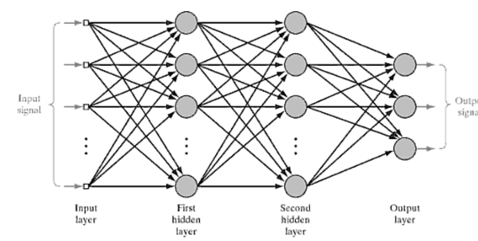


FIGURE 4.1 Architectural graph of a multilayer perceptron with two hidden layers.

solving the supervised learning problem:

error-back propagation for learning weights

Rumelhart, David E.; Hinton, Geoffrey E.; Williams, Ronald J. (8 October 1986). "Learning representations by back-propagating errors". Nature 323 (6088): 533–536

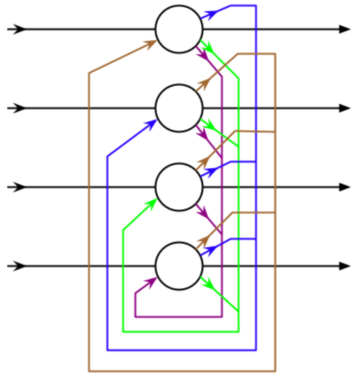
LeCun,Y,Bottou,L,Bengio,Y,andHaffner,P.Gradient-basedlearningappliedtodocumentrecognition. Proceedings of the IEEE, 86(11):2278–2324, November 1998.

*****

Paul J. Werbos. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. PhD thesis, Harvard University, 1974

Bryson, A.E.; W.F. Denham; S.E. Dreyfus. Optimal programming problems with inequality constraints. I: Necessary conditions for extremal solutions. AIAA J. 1, 11 (1963) 2544-2550

1980s



recurrent networks
Hopfield network
Boltzmann machines

- theoretical understanding of what networks were doing
- development of cost (energy) function methods for finding solutions and learning

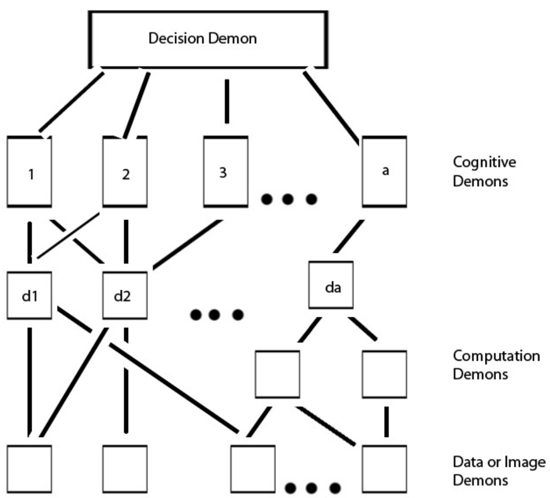
Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science*, *9*(1), 147–169.

The need for an “architecture” for vision

to manage local uncertainty

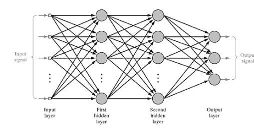
and the complexities of real-world images

Pandemonium 1959

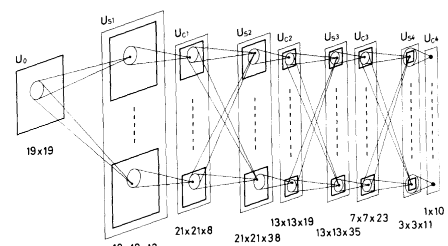


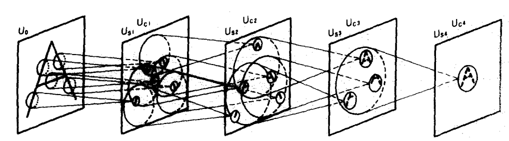
- parallel processing,
- learning
- hill-climbing cost functions

O. G. Selfridge. "Pandemonium: A paradigm for learning." In D. V. Blake and A. M. Uttley, editors, Proceedings of the Symposium on Mechanisation of Thought Processes, pages 511–529, London, 1959.



Fukushima 1988





Fukushima, K. (1988). Neocognitron - a Hierarchical Neural Network Capable of Visual-Pattern Recognition. *Neural Networks*, *1*(2), 119–130.

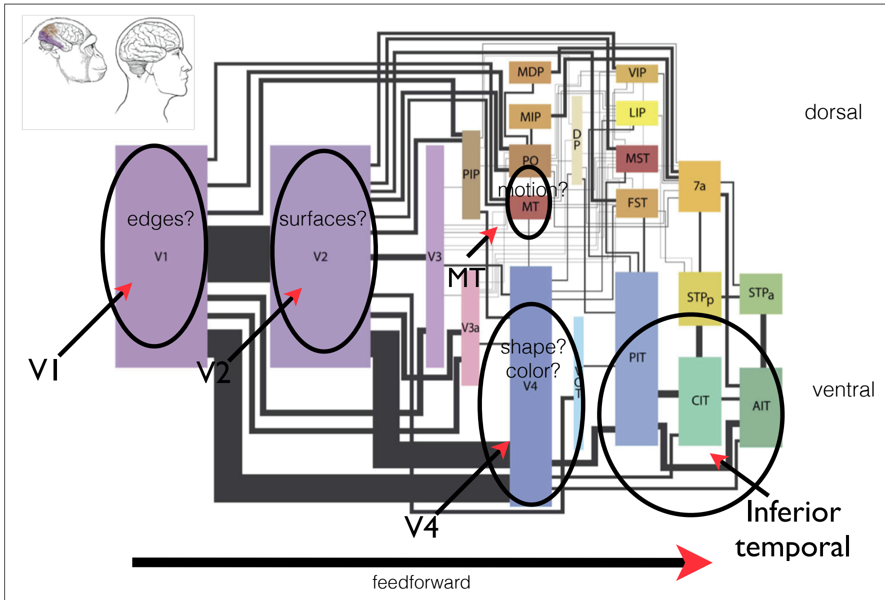
supervised and unsupervised learning

primate visual hierarchical neuroarchitecture

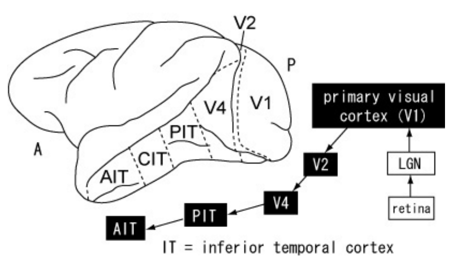
1978….1991

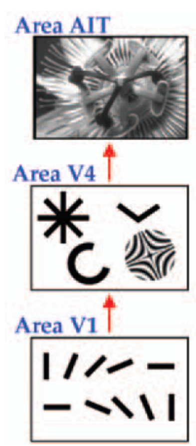
Zeki, S. M. (1978). Functional specialisation in the visual cortex of the rhesus monkey. *Nature*, *274*(5670), 423–428.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cerebral Cortex, 1(1), 1–47.



Hierarchical models of object recognition





Hegde and Felleman, 2007

bread and butter of ventral stream modeling

Hierarchical models for feature extraction for recognition

Local features progressively grouped into more structured representations

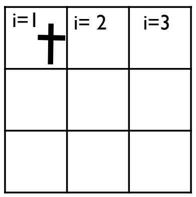
- edges => contours => fragments => parts => objects

Selectivity/invariance trade-off

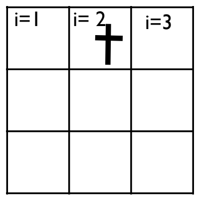
- Increased selectivity for object/pattern type
- Decreased sensitivity to view-dependent variations of translation, scale and illumination

ANDs & ORs
Recognize the letter "t"

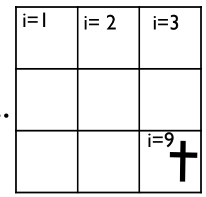
"t" is represented by the conjunction of a vertical and horizontal bar: | AND − = t



OR



OR ...



which can occur at any one of many locations i

"t": h_1 && v_1 || h_2 && v_2 || h_3 && v_3...

---

simple and complex cells as AND- and OR-like operations

contributing towards an end-goal of invariant recognition

Riesenhuber & Poggio model

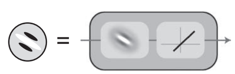
- combine the properties of simple- and complex-like cells with hierarchical organization to achieve invariance

Poggio, T. (2011). The Computational Magic of the Ventral Stream: Towards a Theory. *Nature Precedings*.

---

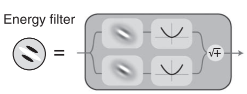
two main classes of V1 cells*

- Simple cells
  - detect conjunctions of inputs
    - similar to a logical AND
    - e.g. of similar pixels to form an edge template
    - "phase sensitive"

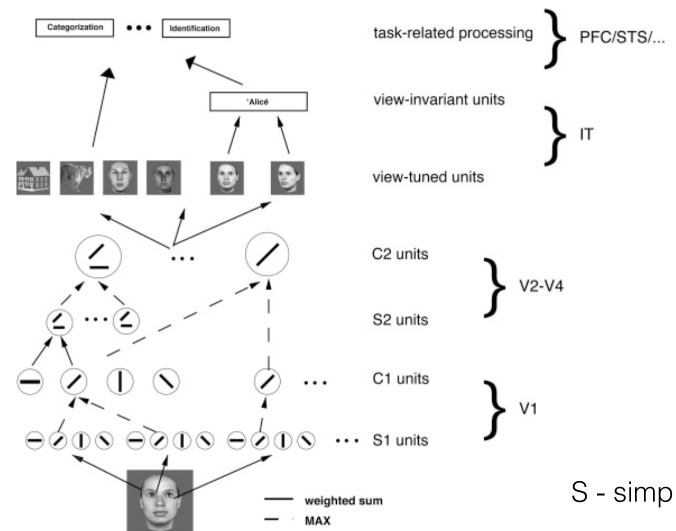


- Complex cells
  - detect disjunctions of inputs
    - similar to a logical OR
    - e.g. any of several similar oriented edges within a region of space will fire cell
    - "phase *in*sensitive"

Energy filter



**The distinction isn't categorical--i.e. a range of phase sensitivities. And there other types of cells, e.g. end-stopped. See Mechler, F., & Ringach, D. L. (2002). On the classification of simple and complex cells. Vision Research, 42(8), 1017–1033.*



S - simple cell like

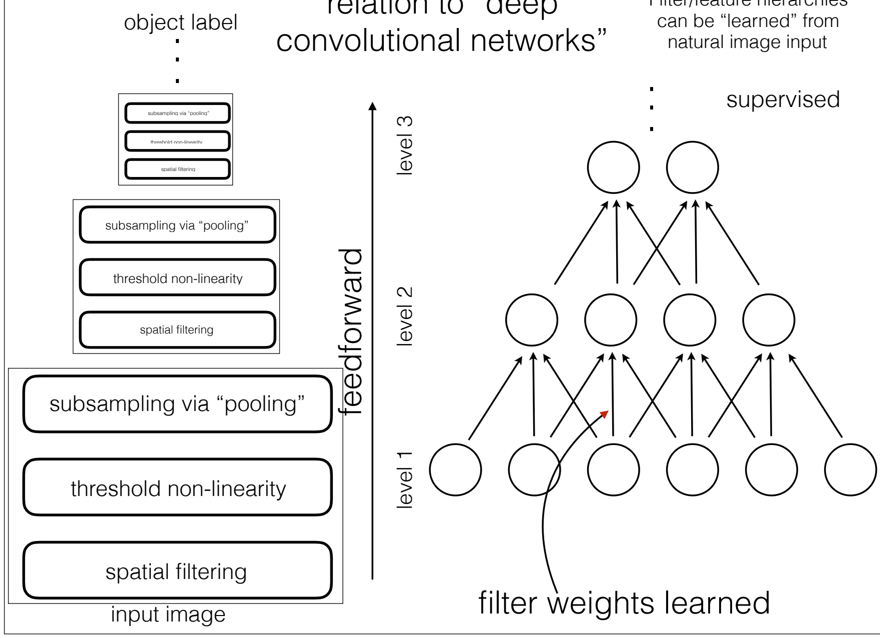
C - complex cell like

Riesenhuber & Poggio, 1999

relation to “deep convolutional networks”

Filter/feature hierarchies can be “learned” from natural image input

supervised



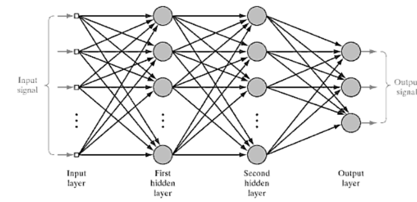


FIGURE 4.1 Architectural graph of a multilayer perceptron with two hidden layers.

Deep convolutional network learning
What’s new since the 1980s?

large labelled image datasets
faster computations—GPUs

What determines feature hierarchies?

Grouping to form more abstract features, given image regularities that support tasks

— “hand - wire” based on analysis of computation and neural models

- e.g. Riesenhuber and Poggio, …

— unsupervised learning based on based on successive discovery of image regularities (Barlow)

- detecting “suspicious coincidences”:
    - Is p(feature A, feature B) >> p(feature A) p(feature B)
    - if so, recode to remove dependence. E.g. contingent adaptation example
- advantage of general features. but perhaps more useful at lower levels of the hierarchy

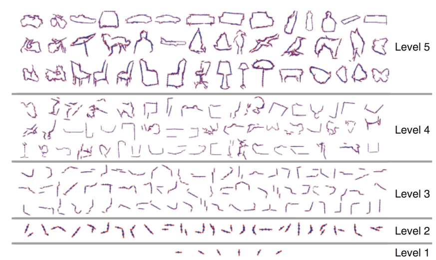
— supervised learning

- — “20 questions” approach (Ephstein et al.)
    - find diagnostic features that distinguish the categories for the most important tasks to determine the top level
    - repeat at a lower level of abstract to find sub-features that distinguish the diagnostic features
    - …and so forth
- deep convolutional networks

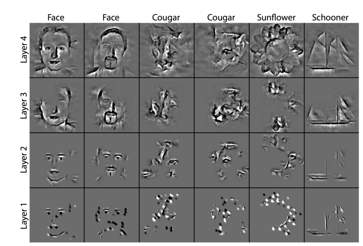
---

unsupervised

Filter/feature hierarchies can be “learned” from natural image input



Zhu, L., Chen, Y., Torralba, A., Freeman, W., & Yuille, A. (2011). Part and appearance sharing: Recursive compositional models for multi-view multi-object detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1919–1926.



Zeiler, M., Taylor, G., & Fergus, R. (2011). Adaptive deconvolutional networks for mid and high level feature learning. Computer Vision (ICCV), 2011 IEEE International Conference on, 2018- 2025.

“Compositional” constraints:
suspicious coincidences
part-sharing

“Deep belief” networks
learning constrained by generative prediction

Explicit, “symbolic”

Implicit

What determines feature hierarchies? An example based on task requirements

Need features for rapid, accurate generalization, given a visual task requirement.

Find features of "intermediate complexity", i.e. image "fragments", that are most informative for category distinctions

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. Nature Neuroscience

Object recognition in the context of a task requirement

What do these scenes have in common?

"Up" curbs-- requiring a step up

Distinguish from non "up curbs"

...that do not require a step up and require different actions

Learning based on informative fragments for the task

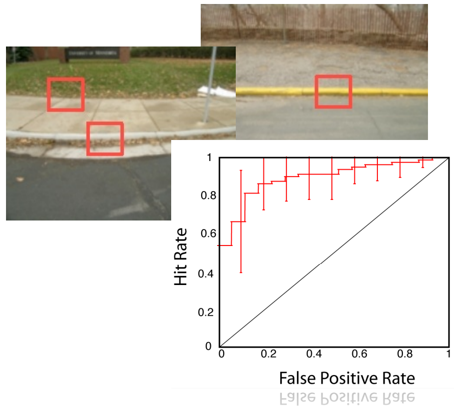
Algorithm finds fragments that maximize mutual information

Detect “up curbs” from an approach angle that requires a step.

View-specific

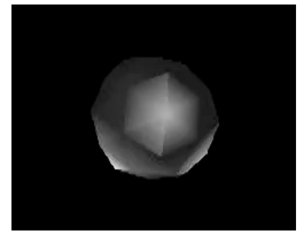
Works well

Experimentally tractable

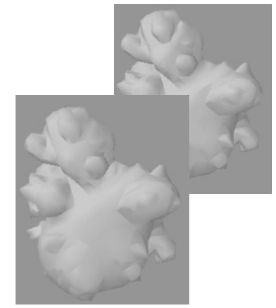


Evgeniy Bart

Do people learn to use fragments of predicted “intermediate complexity”



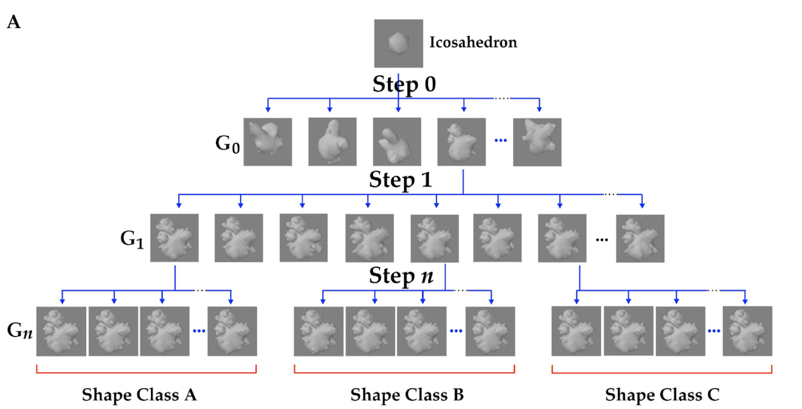
Virtual morphogenesis



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

Generating naturalistic object classes

Virtual Phylogenesis



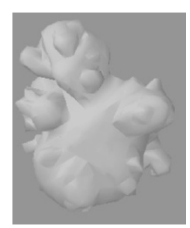
Hegde, J., Bart, E., & Kersten, D. (2008). Fragment-Based Learning of Visual Object Categories. Curr Biol. 18, 597-601

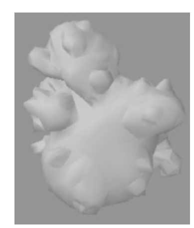
Training

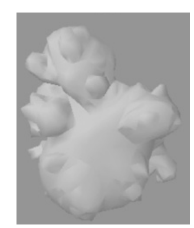
Member of category A or B?

A

B

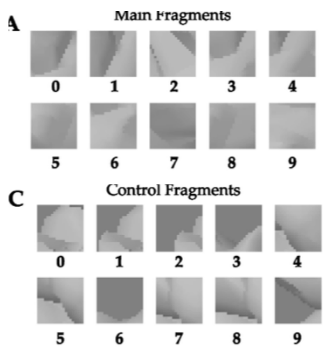


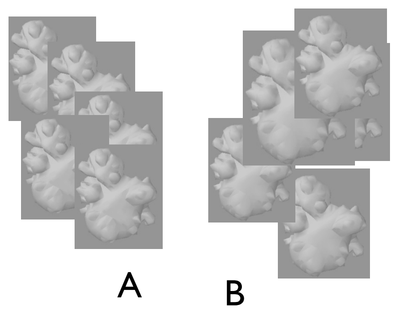




Results

Features of intermediate complexity (local image patches) predicted human observers ability to classify new objects from learned categories





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