Deep learning and human vision

Introduction

artificial neural networks & machine learning

- The first wave: Late 1950s, early 60s:
  - Rosenblatt & the perceptron
- The second wave: mid to late1980s
  - Rumelhart, Hinton &…
- The third wave: early 2010s to present
  - Hinton, LeCun, ….

relevance to neuroscience, psychology, computation?

- too simple to explain dynamics of neural micro-circuitry
  - …but perhaps relevant to larger scale functional architecture?
- descriptive theories of visual behaviorl, lab examples. predictive theories?
- mainly toy applications in computer science/ computer vision

is the third wave different?

https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html

networks that recognize objects given natural images, out-performing state-of-the-art computer vision systems, and competing with people in limited tasks

what if anything does that tell us about mammalian/human vision?

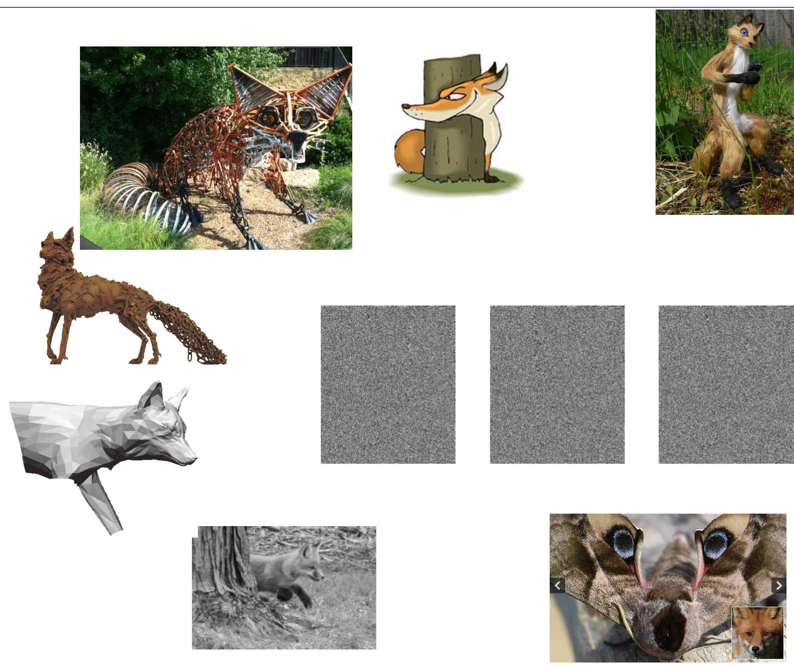
outline

- labeling: a key problem of visual recognition
- shallow models
- deep models
- learning the models
- discriminative vs. generative models
  feedforward vs. feedback

Task: find and name the object category

recognition and the invariance problem

enormous range of appearance variations



discounting/invariance


- Within subordinate-level category
  - e.g. piece of clothing under different lighting, viewpoint, articulation

- Within basic-level category
  - e.g. different types of dogs, coloring and shape details differ, but basic structural appearance is similar


- Within super-ordinate category
  - two reptiles (snake, lizard) can have very different visual appearances

shallow models

image-based models

no explicit knowledge that objects are 3D

view-dependent, image-based models

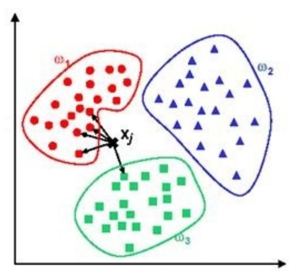
Examples

- Represent each object category by a collection of "snap-shots" of its images or of its "key" 2D features
- Store 2D prototype(s) with model of possible image variations

Examples

Nearest-neighbor

- for a given object, store lots of examples of its images or features, each example has the label for that object
  - represent these in a high-dimensional feature space
- to recognize an object from a new appearance, see what the label is of the nearest stored example



Poggio & Edelman, 1990; Bülthoff & Edelman, 1992; Tarr & Bülthoff, 1995; Liu, Knill & Kersten (1995); Troje & Kersten (1999)

Examples

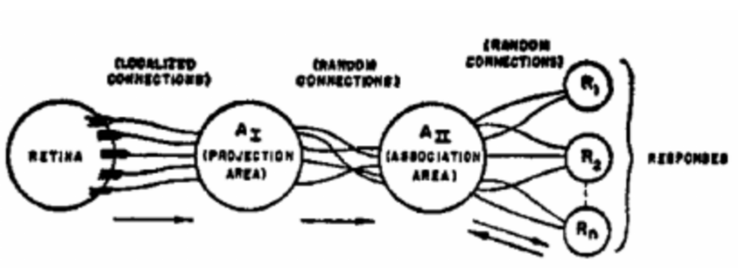
Store 2D prototype(s) with model of possible image variations

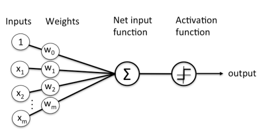
- To recognize new image, either:
  - check to see how close image is to the representation of the prototype (bottom-up/feedforward)
  - manipulate object parameters in memory to check for a match to incoming images/features (top-down/feedback)

behavioral evidence in humans for both kinds of recognition

towards deep models

Rosenblatt's model (1957)





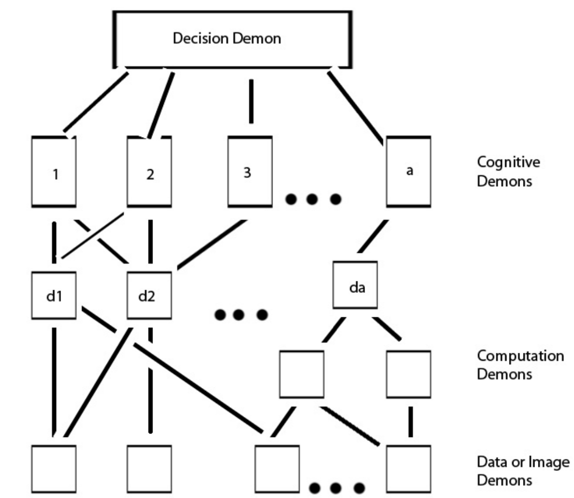
Schematic of Rosenblatt's perceptron.

how deep?

a learning component too, but we'll come back to that

The Perceptron, A Perceiving and Recognizing Automaton, Project Para Report No. 85-460-1, Cornell Aeronautical Laboratory (CAL), Jan. 1957

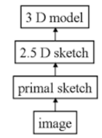
Pandemonium 1959



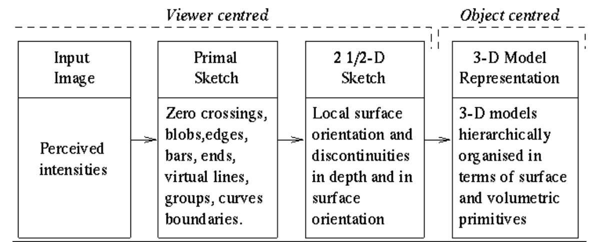
Pandemonium architectures tremendous power because it is capable of recognizing a stimulus despite its changes in size, style and other transformations; without the presumption of an unlimited pattern memory

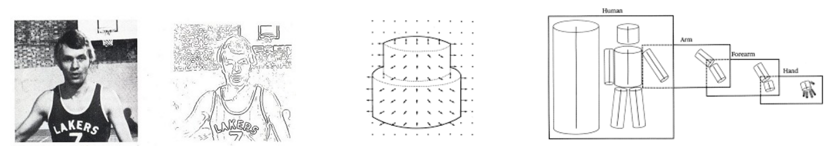
O. G. Selfridge. "Pandemonium: A paradigm for learning." In D. V. Blake and A. M. Uttley, editors, Proceedings of the Symposium on Mechanisation of Thought Processes, pages 511–529, London, 1959.

deeper, hierarchical models



David Marr's 1982 model





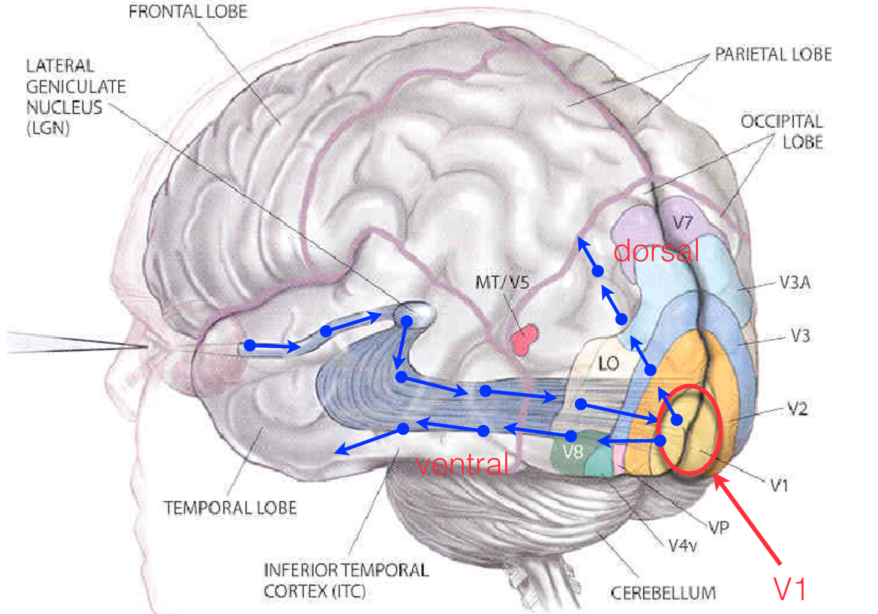
relation to perceptual studies?

- Early vision
  - local image measurements (features) that don't require explicit object knowledge
- Intermediate-level vision
  - grouping of local measures that don't require explicit knowledge of object categories. Only "generic"knowledge
    - symmetry, cue integration, ...
- High-level vision
  - "jobs of vision"
    - compute within-object relations, object-object relations, viewer-object relations

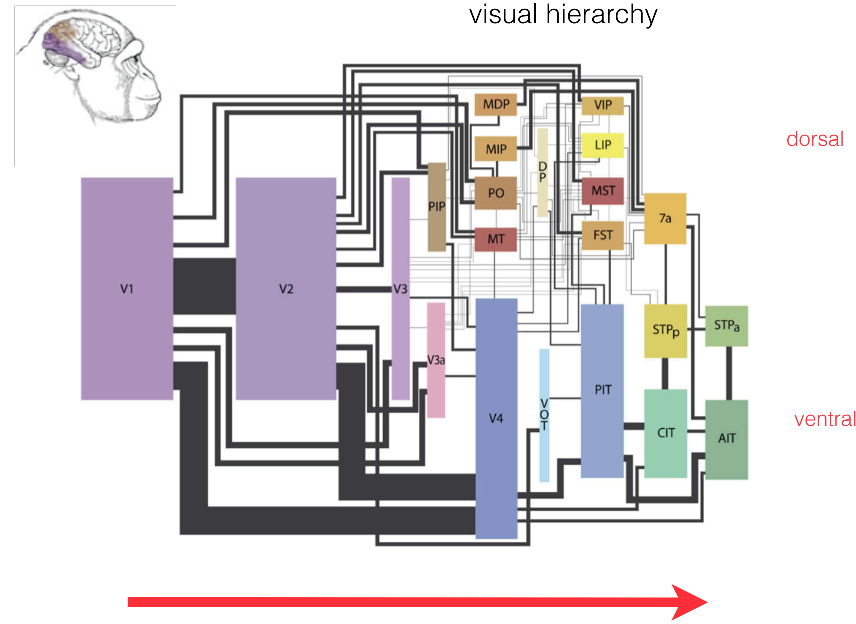
relation to the biology?

- global, hierarchical organization
- local neural circuitry building blocks

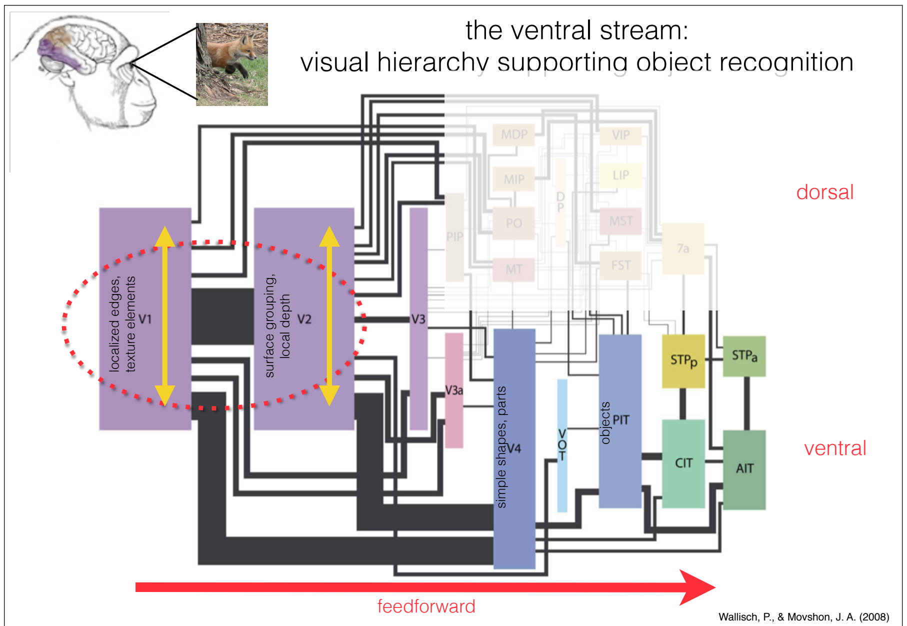
ventral and dorsal visual streams



visual hierarchy



the ventral stream: visual hierarchy supporting object recognition



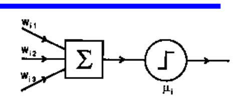
Wallisch, P., & Movshon, J. A. (2008)

local building blocks

some history — 1940s

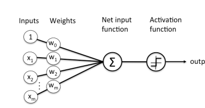
McCulloch–Pitts "neuron"

- Attributes
  - Binary inputs and outputs (0 or 1)
  - Inhibitory inputs are absolute
  - Inputs all have same fixed weight
  - Time invariant
    - Time is quantized in units of synaptic delay



$$n_i(t+1) = \Theta\left[\sum_j w_{ij} n_j(t) - \mu_i\right]$$

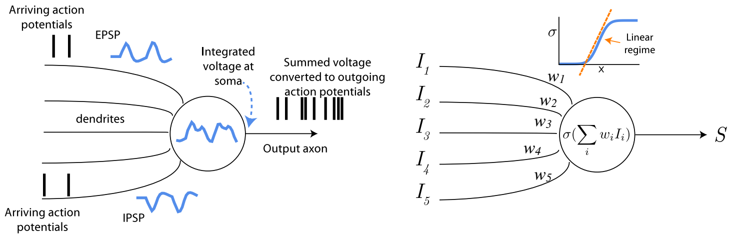
$n_i \equiv$ output of unit i
$\Theta \equiv$ step function
$w_{ij} =$ weight from unit i to j
$\mu_i =$ threshold



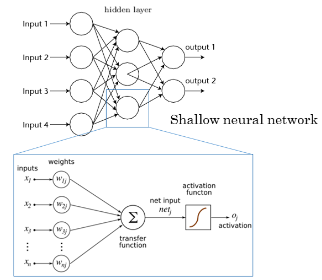
Schematic of Rosenblatt's perceptron.

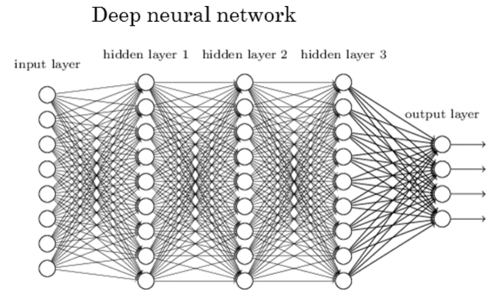
- Key results
  - A synchronous assembly of neurons is Boolean complete
  - The key to computation is the network, not the neuron

local building blocks

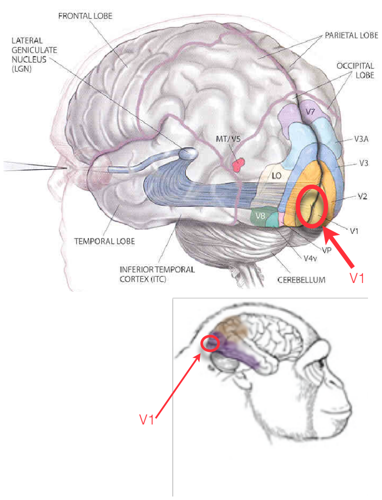


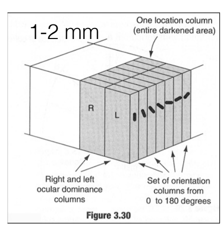
continuous valued inputs and outputs representing frequency of action potentials (spikes"



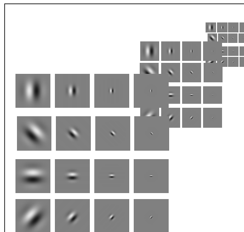


what determines the weights w_i?,

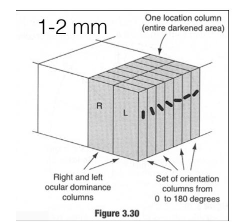




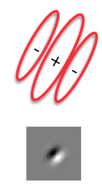
Hubel & Wiesel, 1960s



standard
non-linear spatial filter
V1 model



Receptive field



linear

Normalization

/

Neighboring neurons

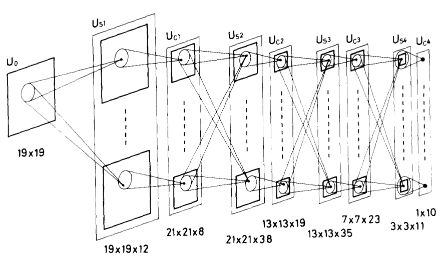
Sparse coding

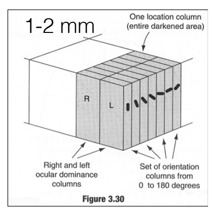
what determines the weights w_{ij} as one proceeds up levels (j) of the hierarchy?,

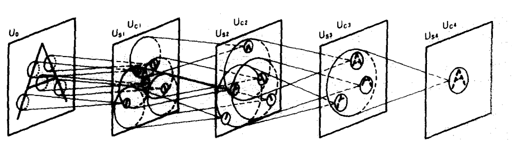
hierarchical models for feature extraction

- 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

Fukushima 1988



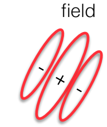




Fukushima, K. (1988). Neocognitron - a Hierarchical Neural Network Capable of Visual-Pattern Recognition. *Neural Networks, 1*(2), 119–130.

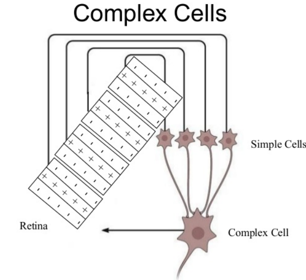
simple and complex cells in V1

simple cell receptive field



one model illustrating local translation invariance

Complex Cells



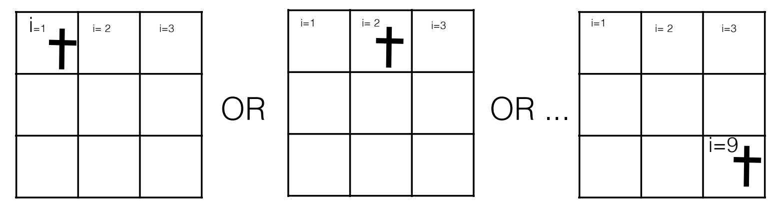
simple & complex cells in V1

- Simple cells
  - “template matching”, i.e. detect conjunctions, logical “AND”
- Complex cells
  - insensitivity to small changes in position, detect disjunctions, logical “OR”
- Recognition as the hierarchical detection of “disjunctions of conjunctions”

Recognize the letter “t”

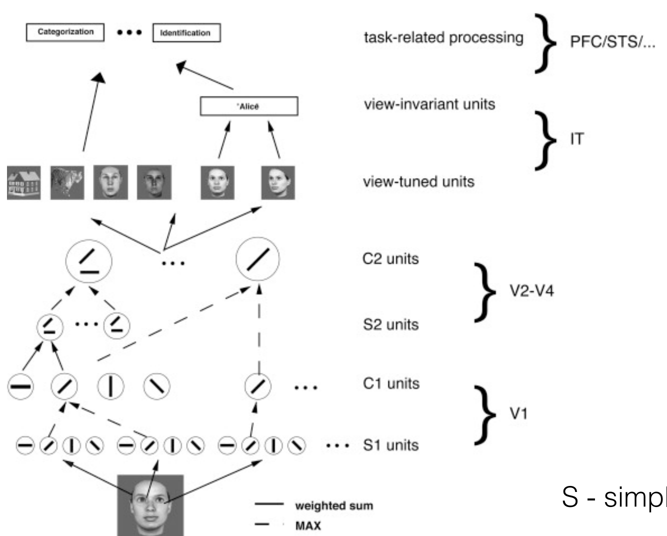
“t” is represented by the conjunction of a vertical and horizontal bar

| AND — = t



which can occur at any one of many locations i

“t”: $h_1 \; \&\& \; v_1 \; || \; h_2 \; \&\& \; v_2 \; || \; h_3 \; \&\& \; v_3 ...$



S - simple cell like

C - complex cell like

Riesenhuber & Poggio, 1999

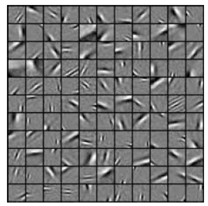
learning the weights?

- instead of “hand wiring”, can the weights be learned?
  - “machine learning”
- two approaches
  - unsupervised learning
  - supervised learning

shallow unsupervised learning

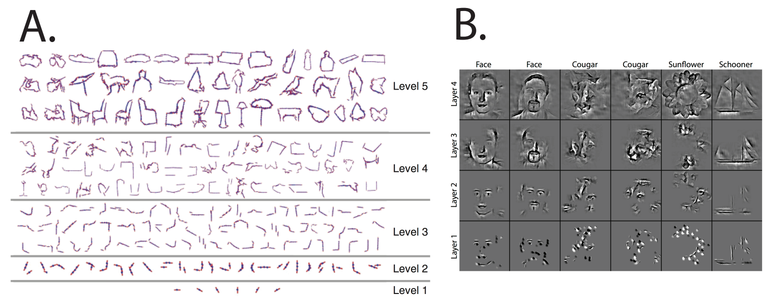
- efficiency constraints, e.g.
  - redundancy reduction
  - sparsity

    find the weights that minimize the number of active V1 model simple cells while preserving the most information about the image
    —Olshausen and Field, 1996



deep unsupervised learning

- find suspicious coincidences, and then recode to eliminate them

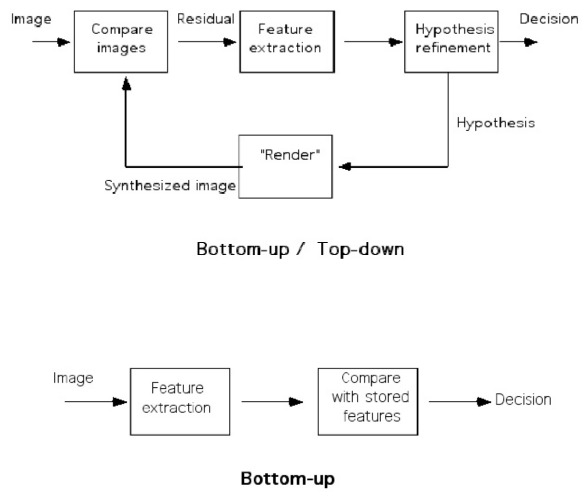


deep supervised learning

- e.g. annotated (i.e. labeled) datasets, with error back-propagation learning
- googlenet

generative vs. discriminative models

feedback and feedforward models



volunteers to lead next week paper discussions?

- Edelman, S. (1997). Computational theories of object recognition. Trends in Cognitive Sciences, 1(8), 296–304.
- DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition Neuron, 73(3), 415–434.