Vision Part 4

Informatics 1 Cognitive Science

Matthias Hennig

School of Informatics
University of Edinburgh
mhennig@inf.ed.ac.uk

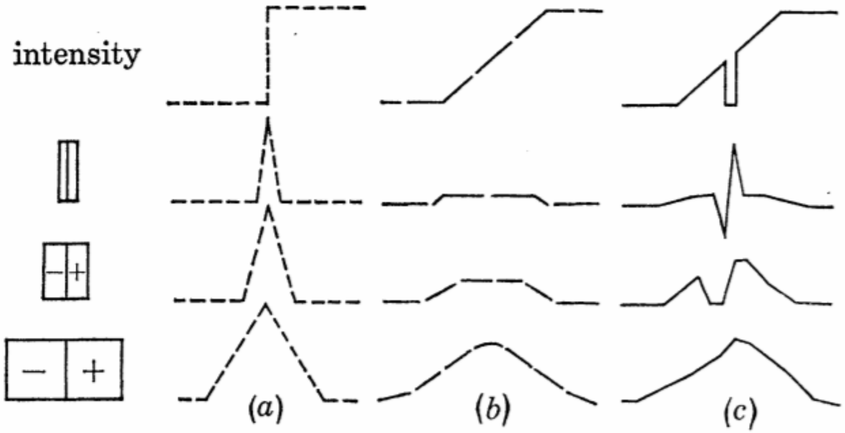
Topics

- Marr and Poggio's Vision Theory
- Object Recognition and the Ventral Pathway
- Deep Neural Networks and the Ventral Stream
- Jennifer Aniston or Grandmother Cells
- Deep learning models, concepts and tricking them

Understanding Vision: Marr & Poggio

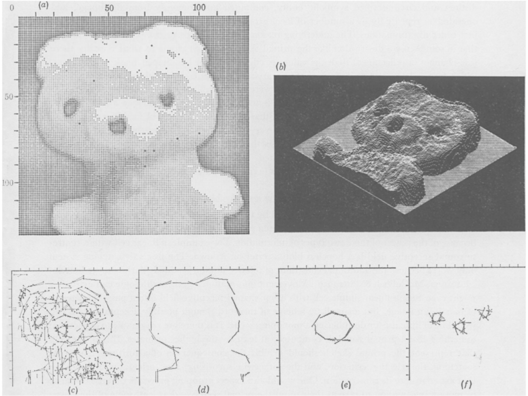
1. Primal sketch: local features including edges, regions, etc.
2. 2.5D sketch: surfaces with depth/orientation — shape as seen by the viewer
3. 3 D model: represents objects in terms of 3D geometric primitives

The Primal Sketch



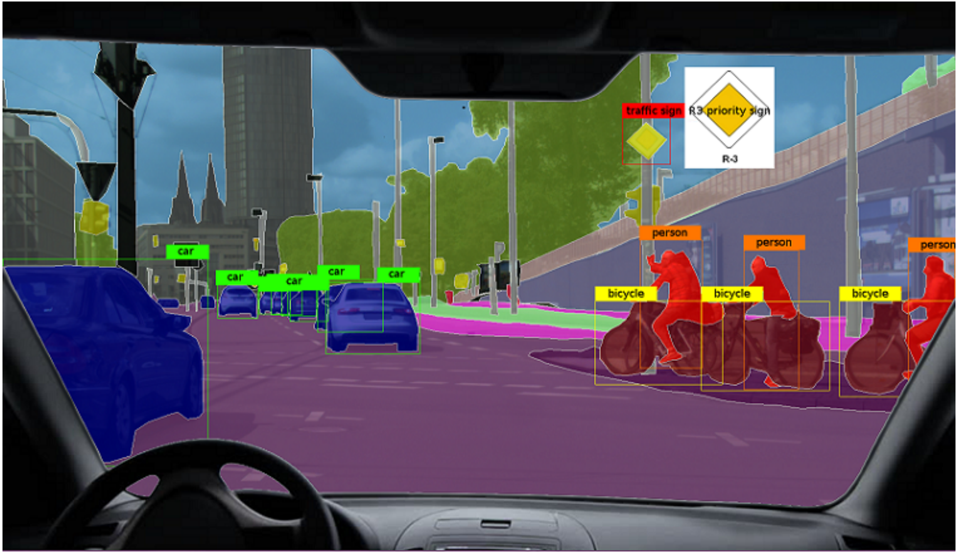
Filters are required to capture contrast changes at different scales. These resemble V1 simple cells RFs.

Beyond the Primal Sketch

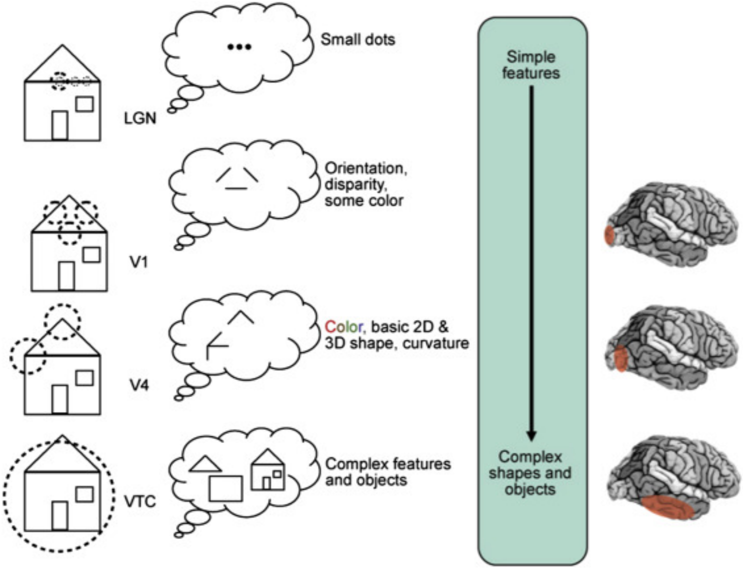


The primal sketch (c), and three principal forms extracted from the primal sketch (d-f). Marr's idea was that these primitives can now be combined to 3D object descriptions. This is a hard problem that deep learning can now solve.

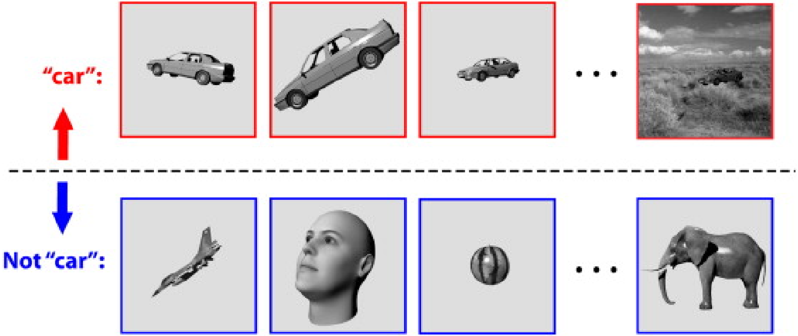
Modern Computer Vision



The hierarchy of visual processing

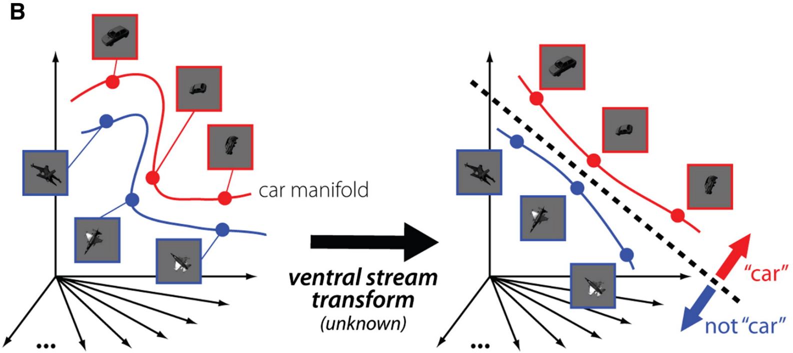

Small dots
LGN
Orientation, disparity, some color
V1
Color, basic 2D & 3D shape, curvature
V4
Complex features and objects
VTC
Simple features
Complex shapes and objects

Object Recognition



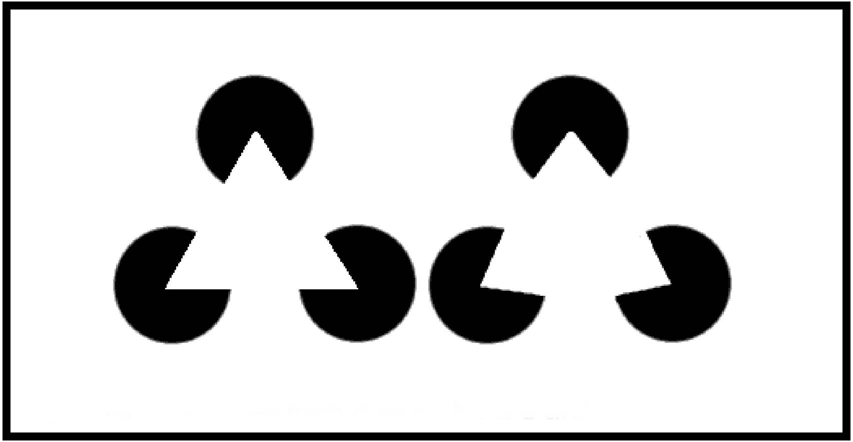
Object recognition is the ability to rapidly (200 ms viewing duration) discriminate a given visual object (e.g., a car, top row) from all other possible visual objects (e.g., bottom row) without any object-specific or location-specific pre-cuing.

Object Recognition



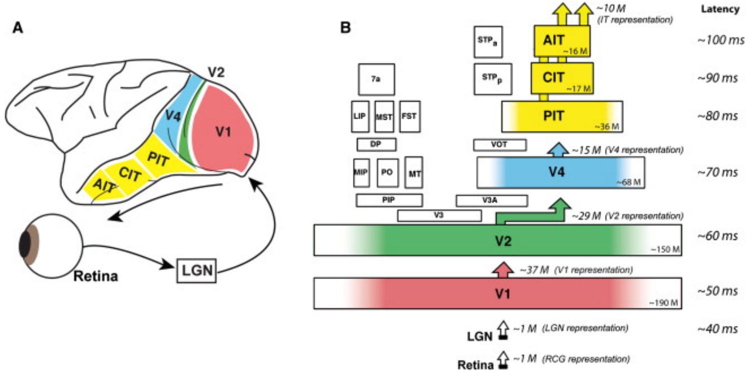
In images and responses in the early visual system, object identity is hidden in curves and tangled "manifolds". The solution is a series of successive re-representations along the ventral stream to a new population representation (area IT) that allows easy separation of one namable object's manifold.

Illusory Contours have a Neural Correlate



Responses corresponding to the non-existing lines in these images are recorded in area V2. This suggests the cortex actively interprets images according to common ecological properties.

The Ventral Pathway

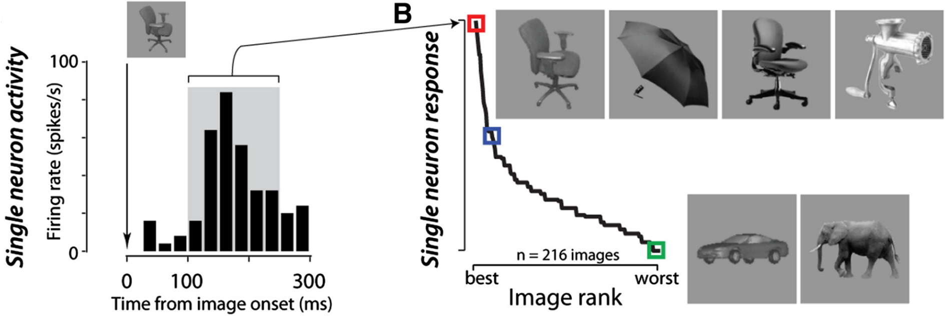


V2: Like V1 and orientation of illusory contours and figure/ground separation

V3: intermediate complexity object features, simple geometric shapes (2.5D-like), but tuning difficult to measure

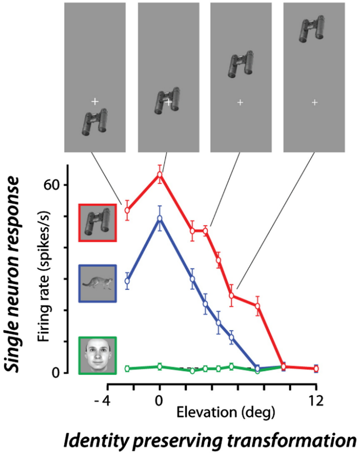
Inferotemporal cortex (IT): complex shapes, objects, and faces

Specificity of IT Neurons



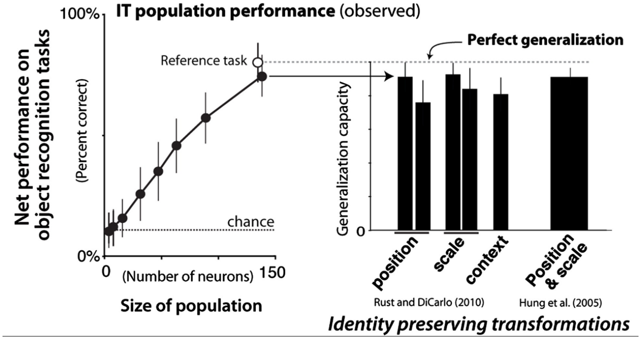
IT neurons respond to pictures of objects with relatively high selectivity. (piatucres from DiCarlo et al., Neuron, 2012)

Invariances in IT Neurons



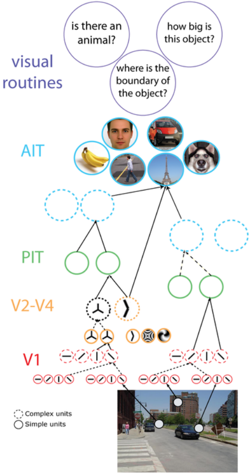
Object preference is preserved over a wide range of elevations.

Decoding Object Identity from IT Neurons



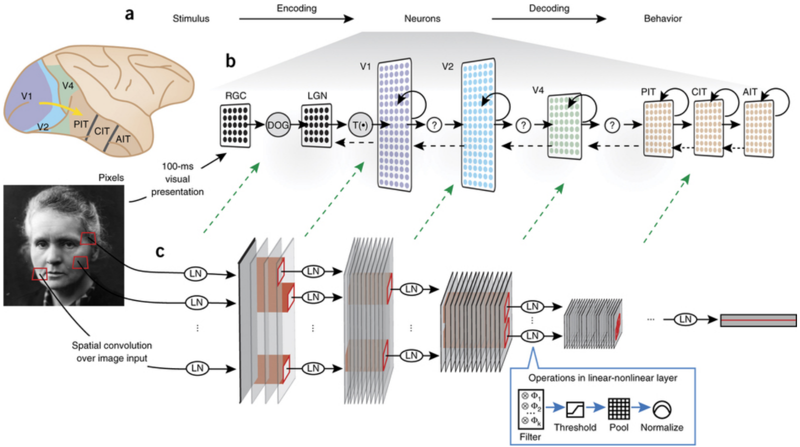
Object classification is near perfect using about 100 IT neurons, and generalisation across position and scale is robust. (reference is a based on SVM classifier on full population)

The HMAX Model - a Model of the Ventral Stream (Riesenhuber & Poggio, 1999)



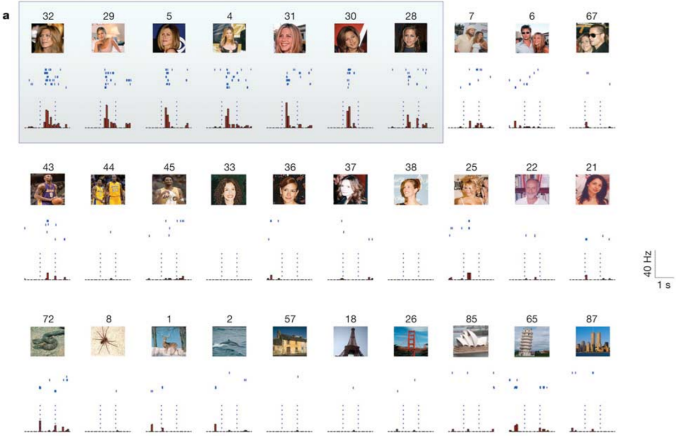
- hierarchical, local layer-wise pooling on multiple scales
- increasing size of RFs
- max pooling in higher layers
- includes learning at the top layer (and intermediate layers in newer version)
- performance ranges 50%-90% in 10 class image data sets

Deep Neural Networks resemble the Ventral Stream



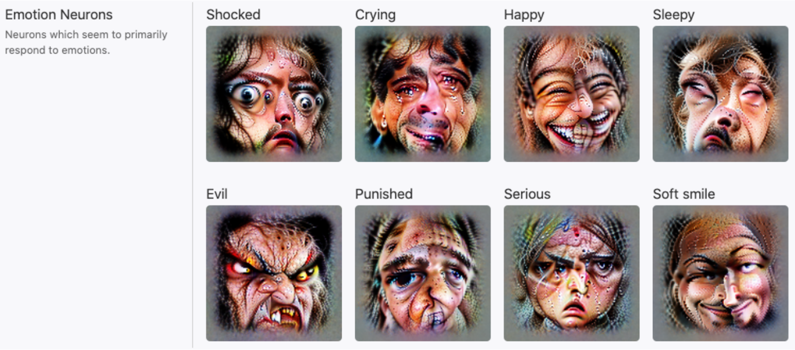
Activations in a deep net trained to classify images mirror recorded activity in the ventral stream, and its hierarchical organisation (Yamis, DiCarlo 2012, 2016).

Jennifer Aniston or Grandmother Cells



A single unit in the hippocampus that responds selectively to images (+ e.g. written or spoken name) of Jennifer Aniston (Quiroga et al., 2005).

CLIP models also have concept cells



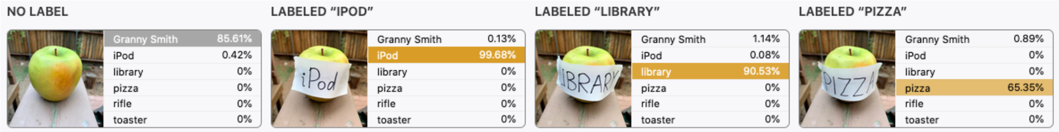
Emotion Neurons

Neurons which seem to primarily respond to emotions.

CLIP model: trained jointly on text and images

Paper: `https://distill.pub/2021/multimodal-neurons/`

CLIP models also have concept cells, but they can be tricked...



NO LABEL

Granny Smith	85.61%
iPod	0.42%
library	0%
pizza	0%
rifle	0%
toaster	0%

LABELED "IPOD"

Granny Smith	0.13%
iPod	99.68%
library	0%
pizza	0%
rifle	0%
toaster	0%

LABELED "LIBRARY"

Granny Smith	1.14%
iPod	0.08%
library	90.53%
pizza	0%
rifle	0%
toaster	0%

LABELED "PIZZA"

Granny Smith	0.89%
iPod	0%
library	0%
pizza	65.35%
rifle	0%
toaster	0%

Stroop effect:

green, blue, red

Summary

- The early visual system is set up to detect local features.
- Along the (in particular ventral) visual pathway, increasingly complex features selectivities are observed.
- Higher visual areas move from features to concepts, objects in images are recognised irrespective of details - this is called invariant object recognition.