Object vision (Chapter 4, part II)

Lecture 9

Jonathan Pillow
Sensation & Perception
(PSY 345 / NEU 325)
Princeton University, Spring 2022
Nonaccidental feature: features that do not depend on the exact (or accidental) viewing position of the observer

**T junctions**: indicate occlusion

**Y junctions**: indicate corners facing the observer

**Arrow Junctions**: corners facing away from observer

• these feature are still present if object is shifted, scaled or rotated by a small amount
viewpoint invariance: the idea that we should be able to recognize an object from any viewpoint.
Problems with view-invariant theories:

Object recognition = not completely viewpoint-invariant!

Viewpoint *does* affect object recognition

- The farther an object is rotated away from a learned view, the longer it takes to recognize

“greebles”  
mental rotations study  
(Gauthier & Tarr 1997)
Face Recognition
Face Recognition: not entirely viewpoint-invariant!
viewpoint invariance (take-away):
- object recognition is *somewhat* but not entirely viewpoint invariant
- observers *do* seem to store certain preferred views of objects.

Makes sense from an evolutionary standpoint: We generate representations that are as invariant as we need them to be for practical applications
Object Recognition

Could a single neuron be responsible for recognizing your grandmother?

“Grandmother cell” - idea of a single neuron responsible for representing some complex object (eg, your granny)

• long considered “idea that could never work”
• how could you have a different neuron for every possible object you know how to recognize?
• what if that neuron died? Could you still recognize your grandmother?
“Jennifer Anniston neuron”

Quiroga et al 2005 (single-cell recordings in humans!)

Inferotemporal (IT) cortex
- high selectivity to people / things, independent of viewpoint
Two facts that constrain any models of object recognition in the visual system
1. Visual processing divided into two cortical streams:

- Separate pathways for “what” and “where” information
Identifying brain regions associated with object recognition:

**Functional imaging** (fMRI) decoding method:

• train a computer to identify images using functional images of brain activity.

• Then examine which brain areas allow for objects to be decoded most accurately
fMRI decoding method:

Training

Image

fMRI scan

Voxel pattern

Output

“BIRD”

Testing

“HAT”

“BIRD?”
Some areas of human cortex are specialized to process certain types of stimuli.

- **Parahippocampal place area (PPA):** responds preferentially to places, such as pictures of houses.
- **Fusiform face area (FFA):** responds to faces more than other objects.
- **Extrastriate body area (EBA):** perception of body parts.
2. Object recognition is fast. (100-200 ms)
Suggests operation of a feed-forward process.

Feed-forward process: computation carried out one neural step after another, without need for feedback from a later stage
(Still debated, but it’s agreed there’s not much time for feedback).
Models of Object Recognition

pandemonium model

• Oliver Selfridge’s (1959) simple model of letter recognition

• Perceptual committee made up of “demons”
  • Demons loosely represent neurons
  • Each level is a different brain area

• Pandemonium simulation:

Models of Object Recognition

Decision Demon

Cognitive Demons

Feature Demons

\[ \text{Review of Models of Object Recognition} \]
Models of Object Recognition

Models of Object Recognition

Decision Demon

Cognitive Demons

Feature Demons

A O X R H

I - / \ ( ) \,

[-] A A H T X X O G R 6 Y
Models of Object Recognition

• Hierarchical “constructive” models of perception:

• Explicit description of how parts are combined to form representation of a whole

  Metaphor: “committees” forming consensus from a group of specialized members

  • perception results from the consensus that emerges
modern version: deep neural networks

Last 10-20 years: rapid progress in “deep learning” methods for object recognition & scene understanding
Captioned by Human and by Google’s image-captioning program

Human: “A group of men playing Frisbee in the park.”
Computer model: “A group of young people playing a game of Frisbee.”

Human: “Three different types of pizza on top of a stove.”
Computer: “A pizza sitting on top of a pan on top of a stove.”
Captioned by Human and by Google’s image-captioning program

**Human:** “Elephants of mixed ages standing in a muddy landscape.”

**Computer:** “A herd of elephants walking across a dry grass field.”
Human: “A green monster kite soaring in a sunny sky.”
Computer: “A man flying through the air while riding a snowboard.”
Summary (Chapter 4)

- middle vision
- gestalt rules (grouping, figure-ground)
- illusory contour
- accidental viewpoint
- non-accidental feature
- viewpoint invariance
- what (ventral) / where (dorsal) pathways
- decoding methods
- pandemonium model
- deep neural network models