Object vision
(Chapter 4, part II)

Lecture 9

Jonathan Pillow
Sensation & Perception
(PSY 345 / NEU 325)
Princeton University, Spring 2023
more fun with accidental viewpoints
The Crevasse
The Crevasse

2-minute “making of” video:  https://www.youtube.com/watch?v=3SNYtd0Ayt0&t=2s
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.
Geons

Biederman 1987: “recognition-by-components” model of object recognition

• visual system identifies objects by recognizing component shapes (“geons”) that compose them

(3 of roughly 36 total geon classes)
“Each geon is kind of like a letter in the alphabet. You can make an infinite number of words out of just 26 letters, just as you might be able to make an infinite number of objects out of several dozen geons.”
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: is 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
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.

Dorsal stream ("where" pathway)

Ventral stream ("what" pathway)

V1

Text

Parietal lobe

Temporal lobe

Where pathway

What pathway
A “wiring” diagram for the monkey visual system

Data from Felleman & Van Essen (1991)
area V4

• cells tuned to stimuli such as spirals, pinwheels, concave and convex shapes

• Difficult to know what V4 neurons do / what stimuli drive them best (but not simple spots or bars!)


Inferotemporal Cortex (IT)

Receptive field properties:

• Very large—some cover half the visual field
• Don’t respond well to spots or lines
• Do respond well to stimuli such as hands, faces, or objects
Inferotemporal Cortex (IT)

• When IT cortex is lesioned, it leads to agnosias (eg object agnosia, prosopagnosia)

• Agnosia: Failure to recognize objects in spite of the ability to see them.
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:

- reveals good decoding of objects from IT and other temporal lobe areas
- does this mean that is what those brain areas DO?

• reveals good decoding of objects from IT and other temporal lobe areas

• does this mean that is what those brain areas DO?
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
“Jennifer Anniston neuron”

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

Inferotemporal (IT) cortex
- high selectivity to people / things, independent of viewpoint
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:

https://learninglink.oup.com/access/content/wolfe6e-student-resources/wolfe6e-activity-4-5-pandemonium?previousFilter=tag_chapter-04
Models of Object Recognition

Decision Demon

Cognitive Demons

Feature Demons

Letters: A A H T X A O G R Y
Models of Object Recognition

Decision Demon

Cognitive Demons

Feature Demons

R

A A H T X X O G R ɓ Y
Models of Object Recognition

Decision Demon

Cognitive Demons

Feature Demons
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
Deep-learning based models

“task based” or “goal based” approaches:

1) train a deep network to perform a recognition task

- Operations in Linear-Nonlinear Layer
  - \( \Phi_1 \), \( \Phi_2 \), ... \( \Phi_n \)
  - Filter
  - Threshold
  - Pool
  - Normalize

Spatial Convolution over Image Input

Behavioral Tasks
  e.g. Trees vs non-Trees

1. Optimize Model for Task Performance

Predicting Neural Responses in Individual IT Neural Sites.

Constructing a High-Performing Model.

Neural Recordings from IT and V4

层面神经元在识别任务中的预测。

[Yamins et al, PNAS 2014]
Deep-learning based models

“task based” or “goal based” approaches:

1) train a deep network to perform a recognition task
2) regress units in trained network against neural data

Operations in Linear-Nonlinear Layer

Spatial Convolution over Image Input

Filter Threshold Pool Normalize

Behavioral Tasks e.g. Trees vs non-Trees

1. Optimize Model for Task Performance

2. Test Per-Site Neural Predictions

Neural Recordings from IT and V4

[Yamins et al, PNAS 2014]
Deep-learning based models

“task based” or “goal based” approaches:

1) train a deep network to perform a recognition task
2) regress units in trained network against neural data

- recent work emphasizes pre-trained networks (VGG, AlexNet, ResNet)
- use of RNNs / LSTMs / GRUs to capture time-course of responses
- current debate about whether we can ever “understand” V1 (or whether that is even a worthwhile goal)
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.”

Captioned by Human and by Google’s image-captioning program

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.”
Captioned by Human and by Google’s image-captioning program

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
- V4, IT
- fMRI decoding methods
- pandemonium model
- deep neural network models