6.819 / 6.869: Advances in Computer Vision

High-level vision

Instructor: Aude Oliva
Lecture TR 9:30AM – 11:00AM
(Room 34-101)
Predictions:

- **type**: outdoor
- **semantic categories**: picnic_area:0.14, patio:0.12, yard:0.11, veranda:0.11, boardwalk:0.06
- **scene attributes**: natural light, man-made, nohorizon, soothing, foliage, trees, vegetation, warm, open area, leaves
High Computing Visual Engine:
Object recognition

0.5 sec
Face & Object recognition: the ventral pathway
Regions of Interest are common brain sub-regions found in everyone.

- Fusiform Face Area (FFA)
- Lateral occipital complex
- Parahippocampal place area (PPA)
Object recognition

- Categories and Function
- Categorization levels: Basic level and entry level
- Typicality effects
- Canonical Perspective
- Prototype theory
- Object parts segmentation
- Mental Imagery
- Brain regions related to object recognition
Organization of knowledge in the brain: Two theories

1) **A category-based approach** suggests that semantic knowledge is organized according to our categories of the world. For instance, one prominent split is between living and non-living things.

2) **A property-based hypothesis** is that semantic knowledge is organized according to the properties (attributes) of objects. These properties may be visual or functional.
Perception of Object function

Function of an object that an observer can observe directly from its structure

Affordances

Perception of Physical Structure  \(\rightarrow\) Perception of Affordances

Flat surface  Horizontal  Knee-high  (etc.)  \(\rightarrow\) Sittable upon

Categorization

Perception of Physical Structure  \(\rightarrow\) Categorization  \(\rightarrow\) Retrieval of Function

Flat surface  Horizontal  Knee-high  (etc.)  \(\rightarrow\) Chair  \(\rightarrow\) Sittable upon

throwable

sittable-upon
Levels of Object Categorization

Object Categorization: superordinate level (e.g. animal), basic level (e.g. preferred level, a tiger), subordinate level (e.g. a Persian tiger).

Basic level organization: Three levels of Regularities:

- **Shape**: Highest level categories for which members have similar shape
- **Motor action**: Highest level categories for which people interact with its members using similar motor sequences
- **Common attributes**: Highest level categories for which they are significant number of attributes in common between pairs of members.

If plot the degree of shape similarity, the degree of similar motor actions, and the number of common attributes: steep decline between the basic and superordinate levels.

Rosch et al (1976) found that when people are shown pictures of objects, they identify objects at a basic level more quickly than they identified objects at higher or lower levels.

Objects appear to be recognized first at their basic level, and only afterwards they are classified in terms of higher or lowers level categories.

Entry-level (Jolicoeur et al., 1984) corresponds to either basic level or subordinate level depending on the object category and exemplar frequency (a taxi instead of car).
Entry-level categories

- The level of which objects are first categorized perceptually
- Basic level: “bird” is the basic level category for every bird
- **Entry-level**: “Bird” is the entry level category for typical birds, but subordinate names (chicken, ostrich) are the entry level categories for atypical birds
Typicality: How common an item is a member of a category

Typical exemplar: a representation of the average (or central tendency)
The representation of a category vary with experience, so does the “typical” exemplar

For tree experts, the internal structure of the category tree is organized around attributes of weediness and height. Thus, the best exemplars of trees are trees of high heights (and free of weedy characteristics like having weak limbs and being susceptible to disease).

Typicality and central tendency illustrated for different kinds of trees varying in height. If typicality is based on central tendency, then trees of average height should be rated as the most typical. Tree experts, however, rate the tallest trees as most typical.
Canonical Perspective

The “best,” most easily identified view of an object.
(Palmer, Rosch & Chase, 1981)

Slide by Palmer
Canonical Viewpoint: 
$\frac{3}{4}$ view – or side view
View with all the parts seen

Invariance in object recognition

(1) Objects can be decomposed in parts

(2) Each object has a distinct configuration of parts

(3) Different viewpoint of the same object must show the same configuration of parts.
Prototype Theory

• Categories are formed on the basis of characteristics features, which describe the typical model of the category.

• Whereas a defining feature is possessed by every instance of a category (e.g. a dog has 4 legs), a characteristic feature need not be (e.g. the ability to fly is typical to birds, but is not a defining feature of a bird).

• According to prototype theory, because an ostrich cannot fly, it seems less birdlike than a robin which can fly.
Prototype Theory

• According to the prototype view, an object will be classified as an instance of a category if it is sufficiently similar to the prototype.

• **Evidence for Prototype:**

• **Typicality ratings:** how good are robins as an example of birds

• **Production order of exemplars:** Name all the kinds of bird you can think of

• **Time to verify categorical statements:** True or false: a robin is a bird

*Figure 7.3. Schematic of the prototype model. Although many exemplars are seen, only the prototype is stored. The prototype is updated continually to incorporate more experience with new exemplars.*
1 Prototype or many Exemplars?

It may depend on the object class.

**Prototype Model**

Category judgments are made by comparing a new exemplar to the prototype.

**Exemplars Model**

Category judgments are made by comparing a new exemplar to all the old exemplars of a category or to the exemplar that is the most appropriate.
Mental representations

- Functional-equivalence hypothesis: Although we do not construct images that are exactly identical to percept, we construct images that are functionally equivalent to percepts.
Principles of Mental Imagery

• Mental transformations of images and mental movements across images corresponds to transformations and movements across physical objects and space

• Spatial relations among elements of a visual image are analogous to these relations in the physical space (for instance, easier to see the details of a larger image in your mind than a smaller one)

• The same regions of the brain are involved in manipulating mental imagery and visual percepts (like imagining manipulating an object)
These findings are functionally equivalent to what we might expect if the participants had been rotating physical objects in space.
Participants seem to have encoded the map in the form of an image and actually to have scanned that image.
Object Size
What’s the template size?

**Hypothesis 1:**
1 template size
For all objects

**Hypothesis 2:**
- consistent visual size
- *different* for each object

**Hypothesis 3:**
- no consistent visual size at all
Mental imagery: No picture on screen

Canonical size:
Object representations are not size invariant

Palmer, Rosch and Chase, 1981
Konkle & Oliva, 2011
Neural Representation of Objects

Are different patches of cortex more likely to process objects of different real-world sizes?
Different cortical regions for small and big objects

Small Objects

Big Objects

View of the bottom surface of an “inflated” brain

Whole brain analysis (n=12)

Individual brains

Ventral Surface

Konkle & Oliva (2012). Neuron
Object Recognition: Multiple sub-regions

High Computing Visual Engine: Object recognition

0.5 sec
Dynamics of Seeing

Cichy, Pantazis, Oliva
MEG: Time
Every millisecond

fMRI: Space
Each millimeter
Representational Geometry
Nikolaus Kriegeskorte (2008)

Shepard et al., 1980; Kruskal and Wish., 1978; Edelman et al. 1998; Kriegeskorte et al., 2008; Mur et al., 2009; Liu et al., 2013
Representational Geometry

Nikolaus Kriegeskorte (2008)

“RDMs as a hub to relate different representations across sensors and models”
Time-specific fMRI searchlight analysis

A spatially unbiased view of the relations in similarity structure between MEG and fMRI

Cichy, Pantazis, Oliva (2016)
Object recognition
Spatiotemporal maps of correlations between MEG and fMRI

100 msec

Visual areas

Inferior-temporal cortex

100 msec

Visual areas

Parahippocampal cortex
Algorithmic-specific fMRI searchlight analysis

A spatially unbiased view of the relations in similarity structure between deep architectures and fMRI

Voxels within searchlight vs. voxel

Spearman Correlation

Layer 3

Cichy, Khosla, Pantazis, Torralba, Oliva (2016) See also Kaligh-Razavi & Krigeskorte (2014)
Spatiotemporal map of correlations between human brain and model layers

AlexNet

Layer 1
Layer 2
Layer 3
Layer 4
Layer 5
Layer 6
Layer 7
Layer 8

V1
Parietal

Layer 1
Layer 2
Layer 3
Layer 4
Layer 5
Layer 6
Layer 7
Layer 8

savannah
field
lake

...
Spatiotemporal maps of correlations between human brain and CNN layers

Layer 1

Layers 1-2

Layers 2-4

Layers 5-8
High-Powered Machine: Principles

0.5 sec
I. Plasticity

Nothing is lost, everything is transformed

Feeling touch with the “visual” brain

Teng, Cichy, Pantazis, Oliva
II. Growth

Immediate  
Short-term  
Long-term

Hippocampus  
- Dentate-