Visual Scene Understanding

Aude Oliva
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Website: http://cvcl.mit.edu
High-level Scene Representation

I. Long-term Memory Representation
What is the fidelity of stored scene representations and the infrastructure that supports them?

Talia Konkle
Timothy Brady
George Alvarez

II. High-level Neural Representation of Visual Scenes
How is the shape of visual scene represented?

Soojin Park
Michelle Greene
Timothy Brady
Memory Representation

**What we know...**

Standing (1973)
10,000 images
83% Recognition

... people can remember thousands of images

**What we don’t know...**

... what people are remembering for each image?

According to Standing

“Basically, my recollection is that we just separated the pictures into distinct thematic categories: e.g. cars, animals, single-person, 2-people, plants, etc.) Only a few slides were selected which fell into each category, and they were visually distinct.”

Dogs
Playing Cards
Welcome to Massive Memory Experiment

A stream of scenes will be presented on the screen for 3 seconds each.

Your primary task: Remember them ALL!

afterwards you will be tested with...

Completely different kinds of places...

Different instances of the same kind of place...
Welcome to Massive Memory Experiment

A stream of scenes will be presented on the screen for 3 seconds each.

Your other task: Detect exact repeats anywhere in the stream.
Barn
Beach
Cavern
Closet
Countryroad
Greenhouse
Waves
Methods – The Study Stream

128 unique semantic categories of natural images

2912 natural images shown in the stream (3 seconds each, 800 msec ISI)

Number of exemplars per category: 4, 16, or 64 !

N= 24 observers
Methods – The Study Stream

Online Task: **Detect Exact Repeats**

Repeats could be 2 to 1024 back in the stream
Repeats could be from categories with 4, 16, or 64 exemplars
7% of images in the stream were repeats (192 / 2912)
Methods – The Memory Test

Followed by 224 **2-alternative forced choice tests**

<table>
<thead>
<tr>
<th>Novel</th>
<th>Exemplar</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.jpg" alt="Novel Image" /></td>
<td><img src="image2.jpg" alt="Exemplar Image" /></td>
</tr>
<tr>
<td><img src="image3.jpg" alt="Novel Image" /></td>
<td><img src="image4.jpg" alt="Exemplar Image" /></td>
</tr>
</tbody>
</table>

None of the tested categories were n-backed
Test Pairs were always the same for all subjects
Any effect of interference is due to the additional exemplars
Results – Recognition Memory

Replication of Standing (1973)

Percent Correct

100
90
80
70
60
50
40
30
20
10
0

1-novel

4
16
64

Exemplar

96
Detailed Representation
Minor Interference

2% drop with doubling the number of exemplars in memory

Konkle, Brady, Alvarez & Oliva (submitted)
Exemplar

Percent Correct

Highly Detailed Minor Interference
Objects & Scenes
Is it fair to compare?

You can make each test item and foil *arbitrarily hard*

We tried to span the *category* with our exemplars and sampled the test item and foil uniformly
Memory for Scenes and Objects

Similar categorical interference effects for objects and scenes

Konkle, Brady, Alvarez & Oliva (submitted)
I. Conclusion

- High fidelity representation in long term visual memory

- Similar categorical interference effects for scenes and objects

- Objects and scenes are entities represented at a similar level of abstraction in long term storage

- The results suggest that the structure of visual categories is information-theoretic optimal: It maximizes within category similarity & minimize between category similarities

See website with papers and stimuli: http://cvcl.mit.edu/MM
Visual Categories are represented by their shape

How to represent the \textit{shape} of scenes?
II – Neural Representation of Visual Scenes

Soojin Park  Michelle Greene  Timothy Brady

semantic category

Walther et al (2009)

global properties: spatial boundary and content

Spatial Layout

Closed  Open

Content

Natural  Urban

Park et al., submitted
Scenes are spatial entities

A scene is a 3 dimensional entity we act within:
it extends in space, it has a size, boundary, content, layout.

Shape of a scene: Spatial Boundary and Content
Spatial Envelope Representation

A scene is inherently a 3D entity that may be described by properties related to its size (volume) and its content.

(1) **Boundary of the space**
   - Mean depth/Size
   - Openness
   - Perspective …

(2) **Content of the space**
   - Naturalness
   - Roughness
   - Clutter …

Spatial Envelope Representation of Visual Scenes

Spatial Boundary & Content
Orthogonal Properties

NATURAL content ← CLOSED spatial boundary → URBAN content

NATURAL content ← OPEN spatial boundary → URBAN content
Experimental Conditions

Spatial Boundary

Closed

Natural

Urban

Content

Open
Experimental Procedure

1-back task

20 blocks per condition
ROIs localized with Independent localizers

PPA

LOC

Epstein & Kanwisher (1998)
Both PPA and LOC regions classified the 4 groups with ~50% accuracy.

An SVM classifier was trained to classify the four conditions using all blocks but one, and then was tested on the remaining block.
Patterns of Errors

The patterns of errors allows to dissociate multiple levels of structure coexisting within intact images, and test the extent to which a specific property is coded in a certain brain area.

Confusion within the same Spatial Boundary

Confusion within the same Content
Patterns of Errors

- Confusion within the same Spatial Boundary
- Confusion within the same Content

Confusion (%)

PPA | LOC
--- | ---
20  | 25
15  | 20
10  | 15
5   | 10
0   | 5

Brain images indicating PPA and LOC regions.
A dual neural pathway for representing the *shape* of a visual scene

Visual scenes are represented in a distributed and complementary manner by different brain regions sensitive to *spatial boundary* vs *content* of a scene

Park, Brady, Greene & Oliva (submitted)
Thank You

Timothy Brady  Talia Konkle  George Alvarez

Soojin Park  Michelle Greene  Timothy Brady

http://cvcl.mit.edu/MM

Funding: National Science Foundation Career Award IIS-0546262