Mind the gap: comparing multiple models of scene representation in brain and behavior

Iris Groen, PhD        New York University
Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing

Nikolaus Kriegeskorte

Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, Cambridge CB2 7EF, United Kingdom; email: nikolaus.kriegeskorte@mrc-cbu.cam.ac.uk

Neural Networks and Neuroscience-Inspired Computer Vision

David Daniel Cox1,2,3,4 and Thomas Dean4,5

Toward an Integration of Deep Learning and Neuroscience

Adam H. Marblestone*, Greg Wayne2 and Konrad P. Kording3

Visual Object Recognition: Do We (Finally) Know More Now Than We Did?

Isabel Gauthier1 and Michael J. Tarr2
Deep nets and object recognition

Krizhevsky, Sutskever & Hinton, 2012

Yamins et al, 2014

Khaligh-Razavi & Kriegeskorte, 2014
beach hut
sand
beach chair

“beach”
Does deep learning explain scene perception?
Outline

• Scene vs. object perception

• fMRI study 1: objects-in-context

• fMRI study 2: comparing multiple models

• How to move forward?
A scene is a semantically coherent (and often namable) view of a real-world environment comprising background elements and multiple discrete objects arranged in a spatially licensed manner.

*Henderson and Hollingworth (1999)*

Scene perception can be usefully contrasted to object perception: whereas objects are spatially compact entities that one acts upon, scenes are spatially distributed entities that one acts within.

*Epstein (2005)*
Scene recognition without objects

Object information degraded

Object information absent

Oliva, 2005
Scene-selective brain regions

See e.g. Epstein (2014) for a review
Understanding scene perception

Do DNNs predict human behavioral scene perception?

Computational models of scene content

Do DNNs predict human brain responses to scenes?

Scene perception behavior

Neural responses to scenes
Outline

• Scene vs. object perception

• fMRI study 1: objects-in-context

• fMRI study 2: comparing multiple models

• How to move forward?
DNNs and object recognition

Yamins et al, 2014

Khaligh-Razavi & Kriegeskorte, 2014
Study 1: objects-in-context + scenes

48 naturalistic image categories
(3 exemplars per category, 2 sets)

accessories
adults
airplanes
appliances
bags
bathrooms
beaches
beds
bikes
birds
boats
body parts
bugs
butterflies
cars
churches
cityscapes
clothes
deserts
dishes
dolls
factories
farm animals
fire
flowers
food
forests
gyms
houses
kids
kitchens
living rooms
lizards/snakes
masks
motorcycles
mountains
older adults
pets
pools
roads
signs
spiders
sports
suburbs
tools
trains
wild animals
Free arrangement behavioral task

“Group these images according to their similarity”
Free arrangement behavioral task

All pairwise comparisons
Behavioral dissimilarity

King &, Groen &, Steel, Kravitz & Baker (in revision)
Behavior

Correlation distance

Euclidean distance

PPA

OPA

Ventral temporal cortex?

Scene-selective cortex?

King & Groen & Steel, Kravitz & Baker (in revision)
Multi-voxel pattern analysis

1-correlation
Behavior

Ventral temporal cortex

Scene-selective cortex

AlexNet layer 8

EUCLIDEAN DISTANCE

CORRELATION DISTANCE

King & Groen & Steel, Kravitz & Baker (in revision)
King & Groen & Steel, Kravitz & Baker (in revision)
Understanding scene perception

Do DNNs predict human behavioral scene perception?
Yes - layer 8 best

Do DNNs predict human brain responses to scenes?
Yes - layer 5 best

Mismatch!
Why the mismatch?

• Nature of stimuli (mix of objects-in-context and scenes)?

• Multi-arrangement task?

• Alternative models for behavior?
Large-scale scene perception

- Mechanical Turk
- SUN database
- > 300 scene categories

SAME or DIFFERENT CATEGORY?

> 2000 participants
> 5 million trials

Greene et al., 2016 (JEP:General)
Comparing behavior to multiple models

Correlation with scene categorization

Greene et al., 2016 (JEP:General)
Three best models of scene perception

Object Model
- Railing
- Building
- People
- Sky
- Trees
- Bush
- Rocks
- Plant
- Grass
- Fence
- Grass
- Pond
- Dirt

Perceptual Model

Function Model
- >1.4 million trials

Scenes
- Car park
- Beach

Action labels
- Sleeping
- Eating

Convolutional Neural Network (AlexNet; 7th layer)

Greene et al., 2016 (JEP:General)
Study 2

- **Goal**: determine how well the top three models predict fMRI responses and behavioral multi-arrangement task

!! Inherent correlations make it difficult to disambiguate models, even within large sets of naturalistic images

(LeSroart, Stansbury & Gallant, 2015; Malcolm, Groen & Baker, 2016)

→ Minimize model correlations through stimulus selection: Iterative sampling from SUN database until we obtain maximally different predictions

→ Use variance partitioning to identify unique contribution of each model
Stimuli by function
Stimuli by DNN
Stimuli by Objects

- airplane_cabin
- control_tower
- youth_hostel
- volleyball_court
- pump_room
- access_road
- tea_garden
- dolmen
- butte
- woodland
- bamboo_forest
- batting_cage
- badminton_court
- playroom
- bat
- bar
- bindery
- lido_deck
- pilothouse
- pier
- volcano
- stilt_house
- underwater_pool
- apse
- bus_depot
- escalator
- skyscraper
- putting_green
- naval_base
- stadium
Model predictions

Object model

DNN model

Function model

fMRI

Behavior

category 1…30

min

distance

max
Model predictions

Object model

DNN model

Function model

category 1…30

category 1…30

fMRI

PPA

max
distance

min

Behavior

Model predictions
Model correlations with behavior

Correlations

Variance partitioning

Groen, Greene, Baldassano, Fei-Fei, Beck & Baker, 2018, eLife
Model correlations with brain

Correlations

Variance partitioning

Groen, Greene, Baldassano, Fei-Fei, Beck & Baker, 2018, eLife
DNN correlations in scene cortex

Object-trained
(ReferenceNet)

Scene-trained
(Places 205)

Groen, Greene, Baldassano, Fei-Fei, Beck & Baker, 2018, eLife
DNN correlations in scene cortex

Groen, Greene, Baldassano, Fei-Fei, Beck & Baker, 2018, eLife
Where are the functions?

Unique correlations with function model outside scene cortex

LOTC: Action perception
(Lignau & Downing, 2015)

Object-trained
(ReferenceNet)

Scene-trained
(Places 205)

Groen, Greene, Baldassano, Fei-Fei, Beck & Baker, 2018
Understanding scene perception

Do DNNs predict human behavioral scene perception?
- Yes, but hand-labeled functions explain additional variance

Do DNNs predict human brain responses to scenes?
- Yes, but not in all regions, and not much gain with higher layers
Outline

• Scene vs. object perception

• fMRI study 1: objects-in-context

• fMRI study 2: comparing multiple models

• How to move forward?
How can we close the loop?

1. DNNs trained on other tasks than object or scene recognition

2. Other types of DNNs (e.g. RNNs), conceptual models

3. Use simpler and more explicit computational models

Malcolm, Groen & Baker (2016) TICS
Neurophysiologically plausible model of scene statistics

Population response to local contrast

Contrast energy

Spatial fragmentation

Contrast response magnitude

\[ \mu = \text{contrast energy} \]

\[ \sigma = \text{spatial fragmentation} \]

Scholte et al., (2009);
Model predicts behavioral categorization

Man-made/natural categorization task

- 1600 scenes
- Including ‘ambiguous’

Model predicts EEG responses...

.. and decoding of behavior from EEG

“Closed loop”

Computational models of scene content

Scene perception behavior

Neural responses to scenes

Behavioral decision

Spatial fragmentation

natural

man-made

Contrast energy

Spatial fragmentation

Spearman’s ρ

β-coefficient

Contrast energy

Spatial fragmentation

0 100 200 300 400

time (ms)

p < 0.05 (FDR)
Theories of scene processing

Global vs. local

Coarse-to-fine

Hierarchy & Reverse Hierarchy of cell types & cortical areas

Feedback connections add details to explicit "vision with scrutiny"

Feedforward hierarchy underlies implicit processing for initial "vision at a glance"

Wolfe et al., 2011

Hochstein & Ahissar 2002
Dynamic feedback for object recognition

Scene complexity determined by local contrast distribution

Neural activity needed for object detection

Low complexity

Frequency
Contrast strength

High complexity

Frequency
Contrast strength

Feed-forward activity only

Feed-forward + feedback

Groen et al., 2018, PLoS Comp Biol
Does deep learning explain scene perception?

Not yet.
Why not?

- Scene perception entails more than object recognition

- High-level information not captured by object-trained DNNs (e.g. functions), is necessary to fully account for scene perception behavior

- Low-level information captured by image statistics, or early DNN layers, may be important for global scene representation

- DNNs capacity to predict (fMRI) responses in vTC does not automatically extend to predicting human scene perception
Thank you!

Jon Winawer
Noah Benson
Dora Hermes
Stephanie Montenegro
Jing-Yang Zhou
Chris Baker
Brett Bankson
Martin Hebart
Dwight Kravitz
Marcie King
Edward Silson
Victor Lamme
Steven Scholte
Sennay Ghebreab
Sara Jahfari
Noor Seijdel

Chris Baldassano (Columbia U)
Morgan Barense (U Toronto)
Diane Beck (U Illionois)
Tessa Dekker (UCL)
Michelle Greene (Bates College)
Assaf Harel (Wright University)
Fei-Fei Li (Stanford University)
Natalia Petridou & BAIR team (UMC Utrecht)
Kandan Ramakrishnan (MIT)