Going after object recognition performance to discover how the ventral stream works. "Invariance" is crux problem.

James DiCarlo  MD, PhD

Professor of Neuroscience
Head, Department of Brain and Cognitive Sciences
Investigator, The McGovern Institute for Brain Research
Massachusetts Institute of Technology, Cambridge MA, USA
Systems neuroscience: the non human primate model
Systems neuroscience: the non-human primate model

Ventral visual stream

Powerful set of visual features
Systems neuroscience: the non-human primate model

Ventral visual stream

Powerful set of visual features
Understanding the brain and discovering game-changing information processing technology are two sides of the same coin.
The convergence of three fields

When biological brains perform better than computers

- New phenomena
  - psychophysics
  - neuroscience
- Attempt to test/falsify those hypotheses
- Falsifiable hypotheses
- New ideas, algorithm parameters

How the brain works

When computers perform as well as or better than biological brains
Common physical source (object) leads to many images

“identity preserving image variation”

View: position, size, pose, illumination

Clutter, occlusion, illumination

Intraclass

Deformation, articulation

Poggio, Ullman, Grossberg, Edleman, Biederman, etc.
DiCarlo and Cox, *TICS* (2007);
The convergence of three fields

New phenomena
- psychophysics
- neuroscience

How the brain works

New ideas, algorithm parameters
- computer science
Brain-inspired computer algorithms

Examples:
- Hubel & Wiesel (1962)
- Fukushima (1980)
- Perrett & Oram (1993)
- Wallis & Rolls (1997)
- LeCun et al. (1998)
- Riesenhuber & Poggio (1999)
- Serre, Kouh, et al. (2005)

FROM BIOLOGY:
- Hierarchy
- Spatially local filters
- Convolution
- Normalization
- Threshold NL
- Unsupervised learning
-...

Serre, Kouh, Cadieu, Knoblich, Kreiman & Poggio 2005
The convergence of three fields

- psychophysics
- neuroscience
- computer science

How the brain works

Attempt to test/falsify those hypotheses

Falsifiable hypotheses

e.g. HMAX
HMAX successes (~2005)

Serre, Kouh, Cadieu, Knoblich, Kreiman & Poggio 2005
HMAX successes (~2007)

(under limited human viewing conditions)
~2008: But HMAX and other models failed to explain neurons

Representational similarity analysis

Biological ventral stream

Models of ventral stream

Kriegeskorte, Frontiers in Neuroscience (2009)
What went wrong?

how the brain works

New phenomena

psychophysics

neuroscience

computer science

New ideas, algorithm parameters

Stringency of these “Brains vs. Machines” tests was far too weak

Attempt to test/falsify those hypotheses

Falsifiable hypotheses

New ideas, algorithm parameters

How the brain works
~2008: Tests of performance were not stringent enough.

Caltech 101 benchmark

Fifteen training sets:

SLF (~HMAX)

Performance (% correct)

100
80
60
40
20
0

state of the art

Performance (%) Animal vs. Non-animal

100
75
50

VI-like

“HMAX 2.0”
(Serre et al. PNAS 2007)

Humans

Head
Close-body
Medium-body
Far-body

Pinto, Majaj, Barhomi, Salomon, Cox, DiCarlo COSYNE 2010

Human level

- IT population
- V1-like
- HMAX

Performance

pixels
2009: More stringent, but compact tests of “object recognition”

Example object recognition task: “car detection”

Image generation strategy:

3D objects

Natural scenes

Test image

Rendered object

View parameters (position, scale, pose, ...)

2009: Toward more stringent tests of “object recognition”

Basic car task, variation level: 3

“car”

not “car”

2010: Machines vs. human brains

Machines beat humans!

Data merged here: 48 basic-level tasks (8 labels x 6 level of variation)

Pinto, Barhomi, Cox & DiCarlo, WACV(2010)
Human level

- IT population
- V1-like
- HMAX

Performance

pixels
Human level

- IT population
- V1-like
- HMAX
- Pixels

Performance
Human level

simple decode
IT population

V4 population

V1-like

HMAX

pixels

Performance
Neural population similarity of images along the ventral stream

Yamins, Hong, Soloman, Seibert and DiCarlo (under review)

Inspired by N. Kriegeskorte et al. (2008, 2009)

Current maximum expected explanatory power *

Explanatory power of HMO model

HMO model

Animals (8)
Boats (8)
Cars (8)
Chairs (8)
Faces (8)
Fruits (8)
Planes (8)
Tables (8)

Image generalization
Object generalization
Category generalization

Image generalization
Object generalization
Category generalization

Dissimilarity of population similarity to IT

Other models

Spearman correlation coefficient
Predictions of single site IT responses from current best model

Yamins, Hong, Soloman, Seibert and DiCarlo (under review)

Ability to predict IT responses to new images and new objects is dramatically better than previous models.
Basic bio-constrained model component inside HMO

**Basic operations:** \( \Theta = (\theta_{\text{filter}}, \theta_{\text{thr}}, \theta_{\text{sat}}, \theta_{\text{pool}}, \theta_{\text{norm}}) \)

Hierarchical Stacking


"Output" is thousands of visual features

The better a model performs, the better it explains IT responses.

Performance of artificial visual features (% correct)

Ability of artificial visual features to predict IT responses (% variance explained)

Exploration of basic model class

We are optimizing this way

(2013)
Human level

Simple decode

IT population

HMO

Super Vision

Zeiler & Fergus

V4 population

V1-like

HMAX

Pixels

Today:

Performance
Follow the performance trail...

How the brain works

- New phenomena
  - psychophysics
  - neuroscience

- Attempt to test/falsify those hypotheses

- New ideas, algorithm parameters
  - computer science

- Falsifiable hypotheses

Stringency of these tests is crucial. Must include “invariance”.
The power of stringent tests to elucidate biological brains

1) • Discover IT neuronal codes that can explain behavior
   • Demonstrate that other possible codes CANNOT
   • Demonstrate which computer vision features CANNOT

2) • Driving discovery ("learning?") of new CV features
   • These are becoming more and more capable of explaining what the brain is doing