The future of Brain-Score

Martin Schrimpf

msch@mit.edu

@martin_schrimpf

Brain-Score

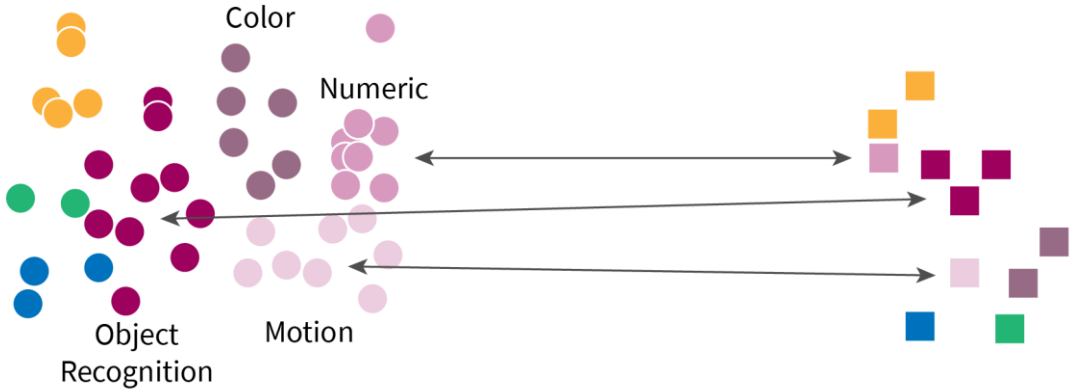
MIT Quest for Intelligence

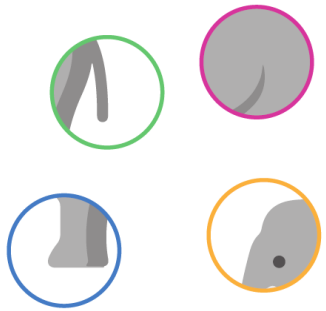
brain+cognitive sciences

McGOVERN INSTITUTE FOR BRAIN RESEARCH

CENTER FOR Brains Minds+ Machines

Piecewise Efforts

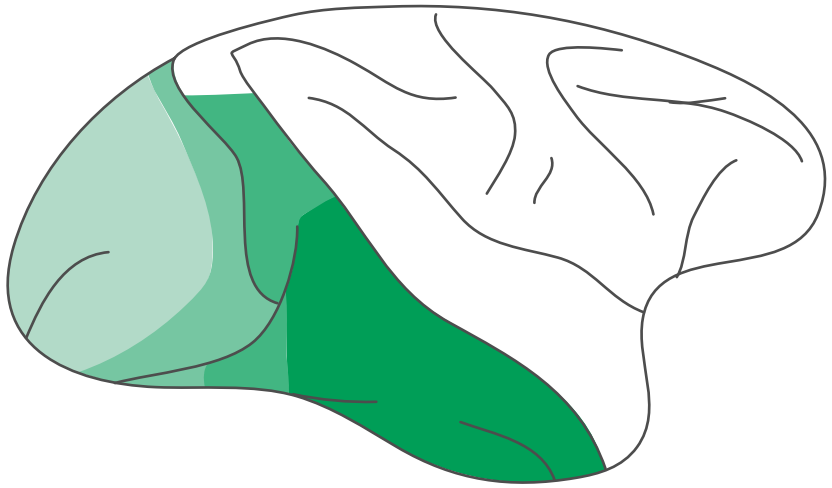


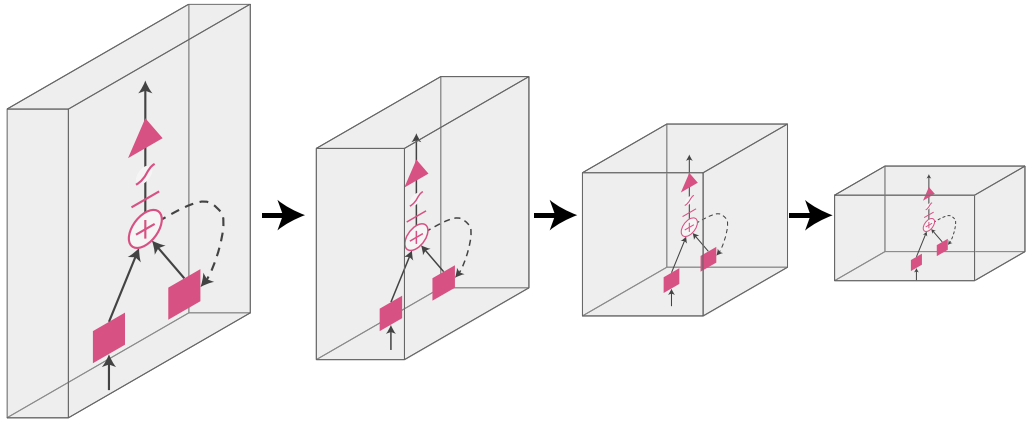


These are necessary first steps! But insufficient for a unified model by themselves

Schrimpf et al. (Neuron 2020)

System Models of Visual Brain Processing



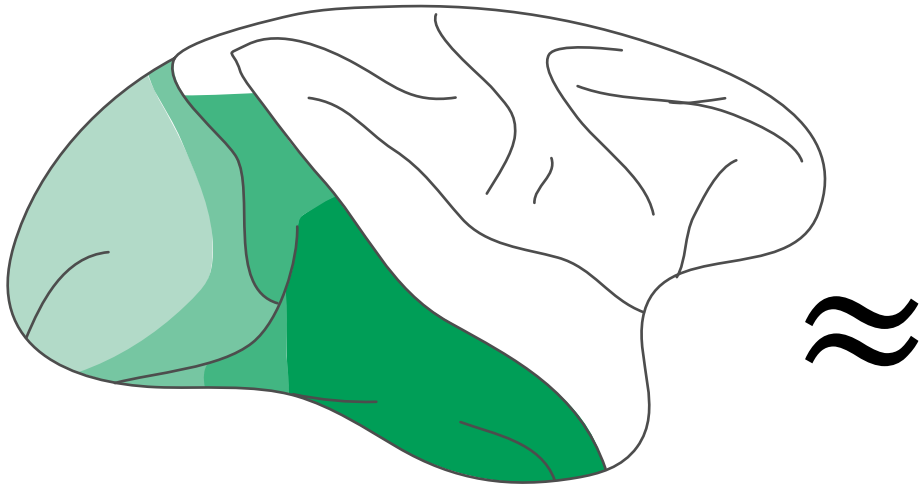


multi-stage: all brain regions

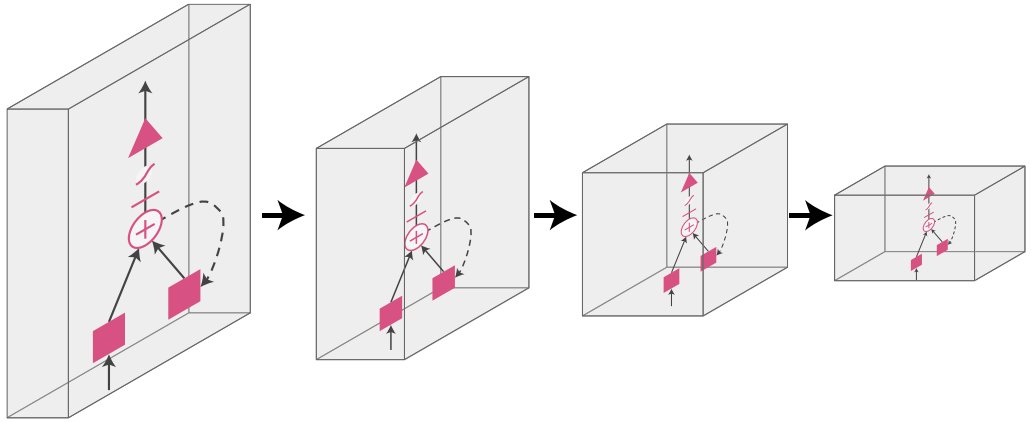
multi-scale: measurements of single neurons to neuronal populations

Schrimpf et al. (Neuron 2020) *Marques, Schrimpf, DiCarlo (bioRxiv 2021)*

Imagine we had System Models of Visual Brain Processing



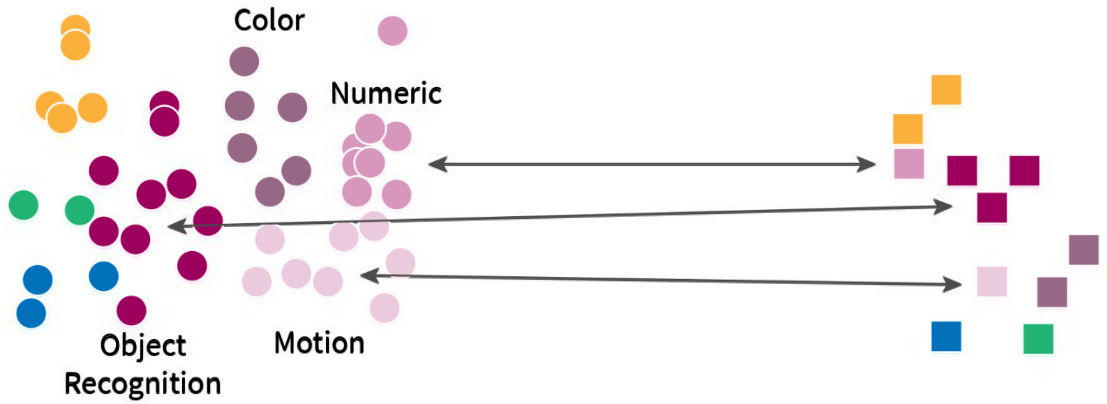
≈



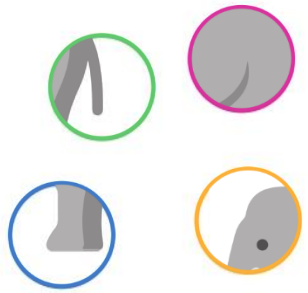
Future clinical applications ← "end-to-end" mechanistic understanding of visual intelligence → Next-generation AI algorithms

Schrimpf et al. (Neuron 2020)

Piecewise Efforts



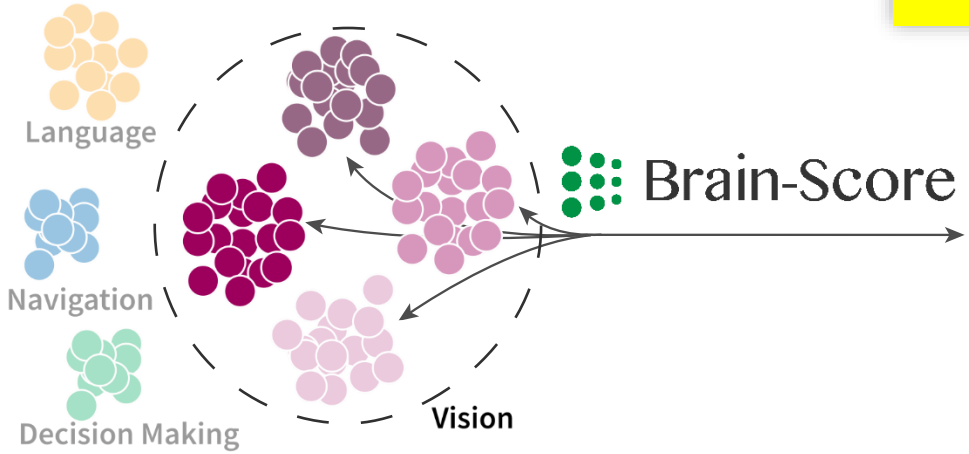
Isolated Data

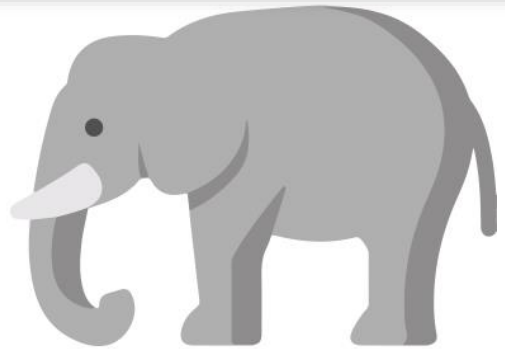


Isolated Models

The only way we can think of to make progress on system models is integrative benchmarking

Integrative Benchmarking





System Models

Schrimpf et al. (Neuron 2020)

More and more domains are beginning to assemble brain benchmarks to test models

Vision

Neuron

Perspective

Integrative Benchmarking to Advance Neurally Mechanistic Models of Human Intelligen

Martin Schrimpf,[1,2,3] Jonas Kubilius,[2,4,5] Michae
and James J. DiCarlo[1,2,3,*]
[1]Department of Brain and Cognitive Sciences, MIT, Car
[2]McGovern Institute for Brain Research, MIT, Cambridg
[3]Center for Brains, Minds and Machines, MIT, Cambrid
[4]Brain and Cognition, KU Leuven, Leuven, Belgium
[5]Three Thirds, Vilnius, Lithuania
*Correspondence: dicarlo@mit.edu
https://doi.org/10.1016/j.neuron.2020.07.040

Multi-scale hierarchical neural ne
the primate primary visual corte

Tiago Marques, Martin Schrimpf, Jar
doi: https://doi.org/10.1101/2021.03.01.4334

Your head is there to move you around: Goal-driven models of the primate dorsal pathway

Patrick J Mineault ...mail.com
Shahab Bakhtiari
Mila, McGill University

Comparing (many) models on brain data is becoming increasingly popular

Audition

Article

A Task-Optimized Neural Network Replicates Human Auditory Behavior, Predicts Brain Responses, and Reveals a Cortical Processing Hierarchy

Alexander J.E. Kell [1, 2, 6, 7], Daniel L.K. Yamins [3, 4, 6], Erica N
McDermott [1, 2, 5]

Deep neural network models reveal interplay of peripheral coding and stimulus statistics in pitch perception

Mark R. Saddler, Ray Gonzalez & Josh H. McDermott

Assembling (many) brain data into quantitative benchmarks will be widely useful

Language

The neural architecture of language: Integrative modeling converges on predictive processing

Martin Schrimpf[a,b,c,1], Idan Asher Blank[a,d,2], Greta Tuckute[a,b,2], Carina Kauf[a,b,2], Eghbal A. Hosseini[a,b], Nancy Kanwisher[a,b,c,1], Joshua B. Tenenbaum[a,c,3], and Evelina Fedorenk

Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain)

Language proce
computational c

Charlotte Caucheteux

Shared compu
processing in

Ariel Goldstein, Zaid Zada, ... Uri Hasson + Show authors

Mariya Toneva Leila Wehbe
Does the brain represent words?
An evaluation of brain decoding studies of language understanding

Jon Gauthier (jon@gauthiers.net) and Anna Ivanova[1] (annaiv@mit.edu)

(not every domain is ready for this)

"Nev

III-060

A high-throughput pipeline for evaluating recurrent neural networks on multiple datasets

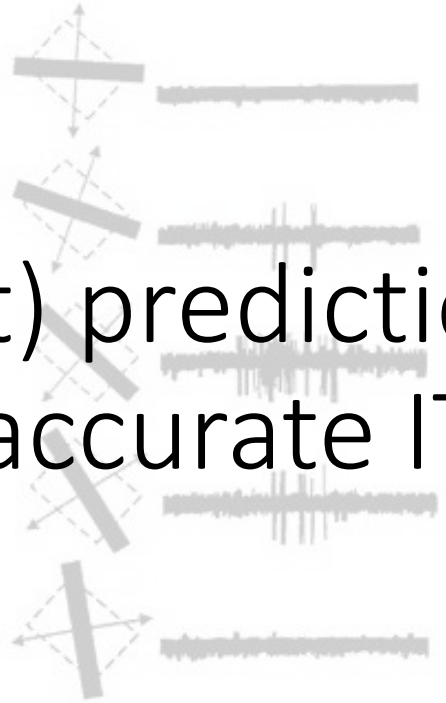
Li, Nathan Cloos, Xun Yuan, Guangyu Robert Yang, Christopher J. Cueva

ural network modeling of a visually-guided social behavior

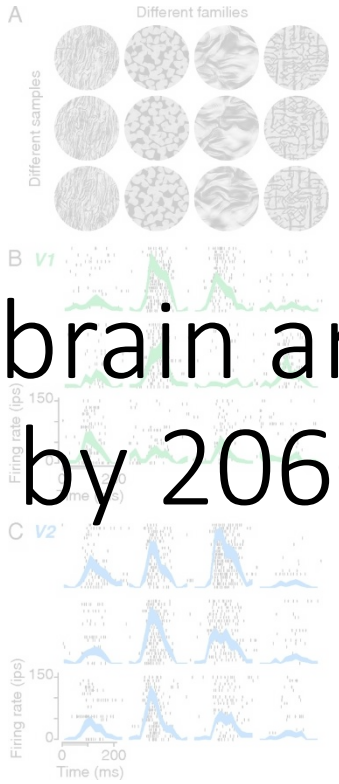
Cowley, Adam Calhoun, Nivedita Rangarajan, Jonathan Pillow, Mala Murthy

Eero's (pre-deepnet) prediction: one brain area every 10-20 years, accurate IT model by 2060

Enroth-Cugel et al., 1966



Hubel and Wiesel, 1968



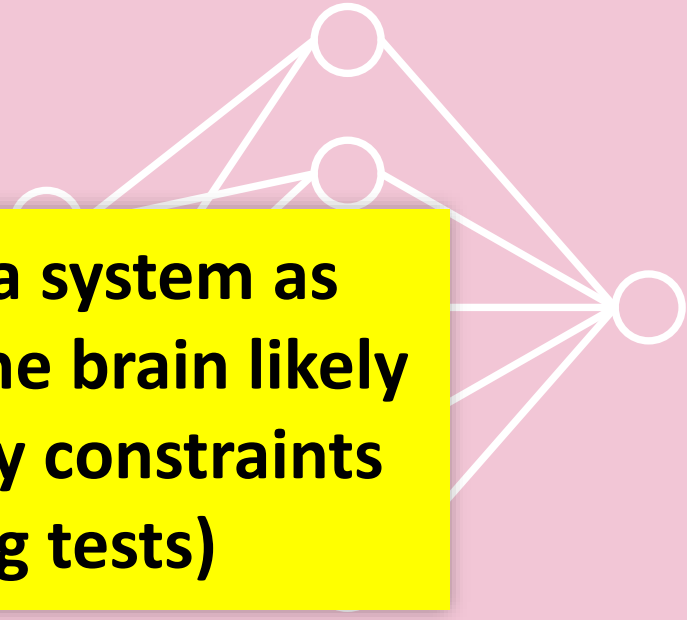
Freeman*, Ziemba*, et al. 2013

The more efficient we can make the cycle of hypothesis creation and testing, the faster progress we will make

Brain and Cognitive Sciences

Computational Models

Sharing quantitative *benchmarks* makes brain modeling accessible and reproducible



Models of a system as complex as the brain likely require many constraints (guiding tests)

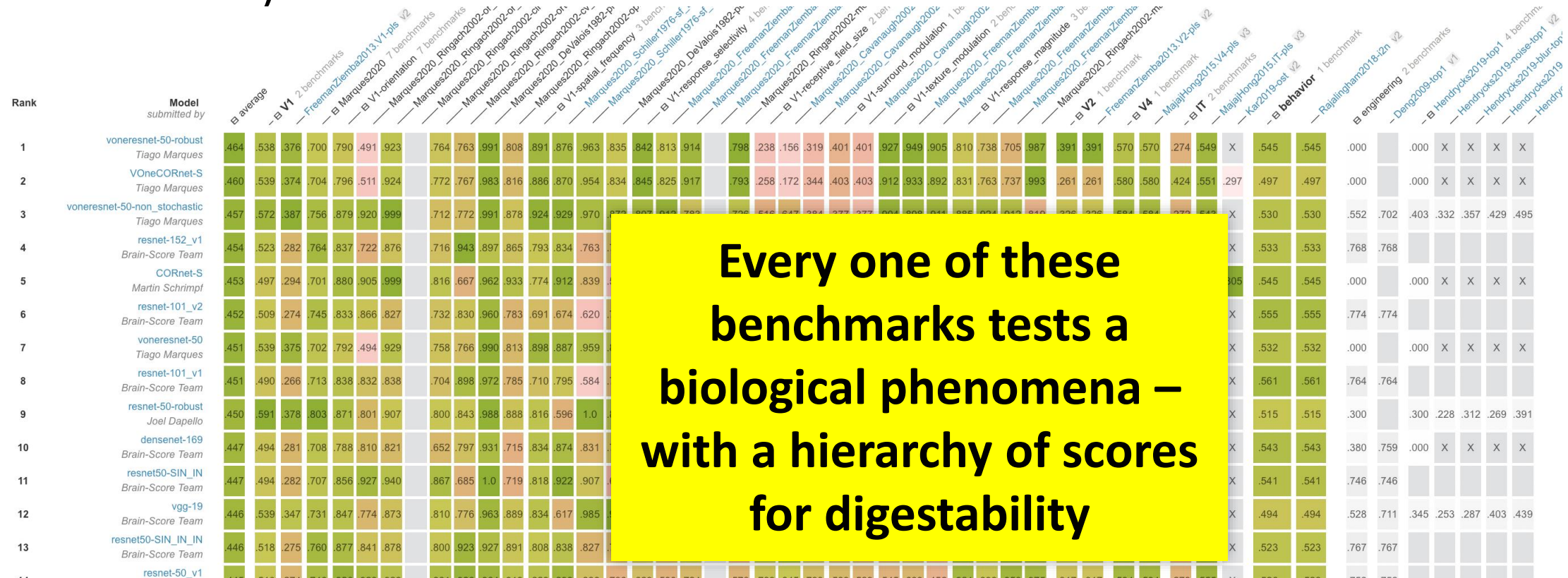
Health, education

Computational understanding of natural intelligence

AI

Brain-Score is a growing ecosystem for testing brain models of vision

- 34 benchmarks spanning V1, V2, V4, IT, behavior (+ engineering benchmarks)



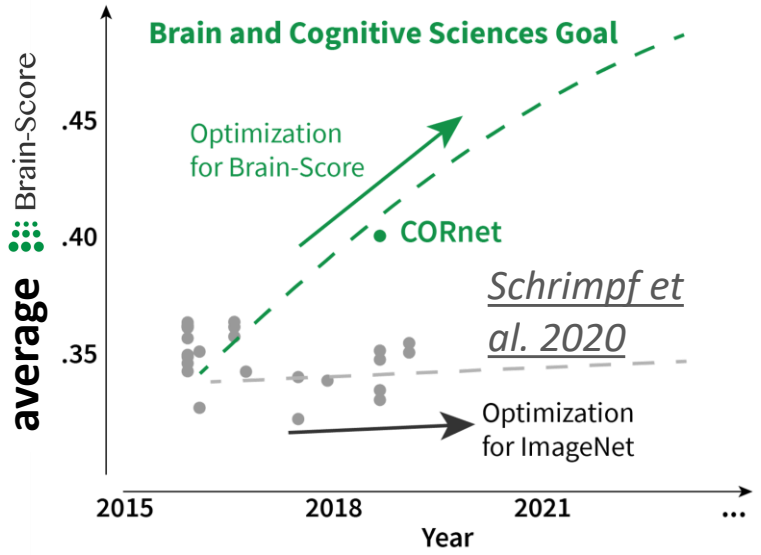
Brain-Score is a growing ecosystem for testing brain models of vision

- 34 benchmarks spanning V1, V2, V4, IT, behavior

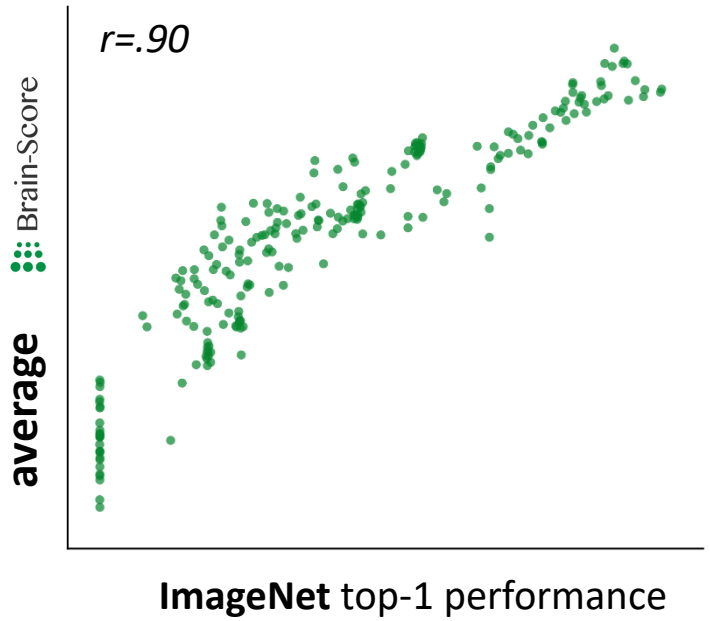
- Top model scores (under current benchmarks)
  - 50% average
  - 60% V1
  - 40% V2
  - 49% IT
  - 66% behavior

Brain-Score enables the discovery of relationships

1 **Building ever more accurate unified ventral stream models**



Brain-Score enables the discovery of relationships



Neural Mechanisms

Computational Task

Behavior

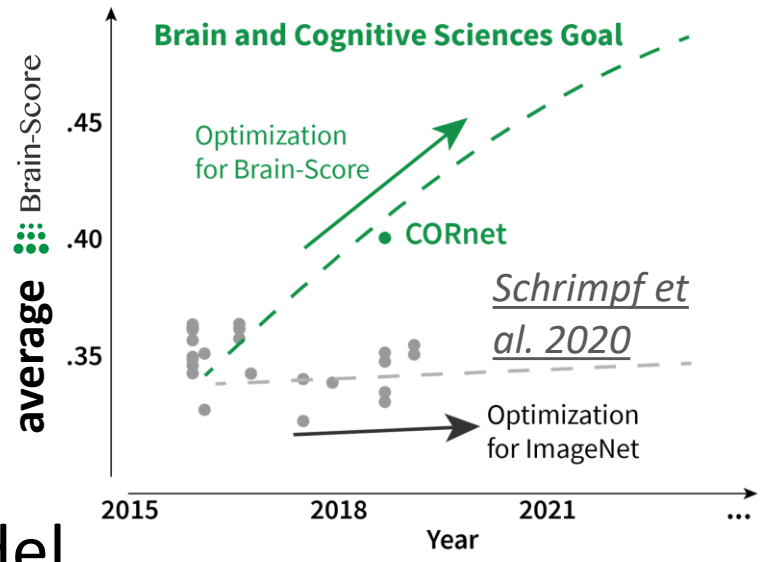
Brain-Score enables the discovery of relationships

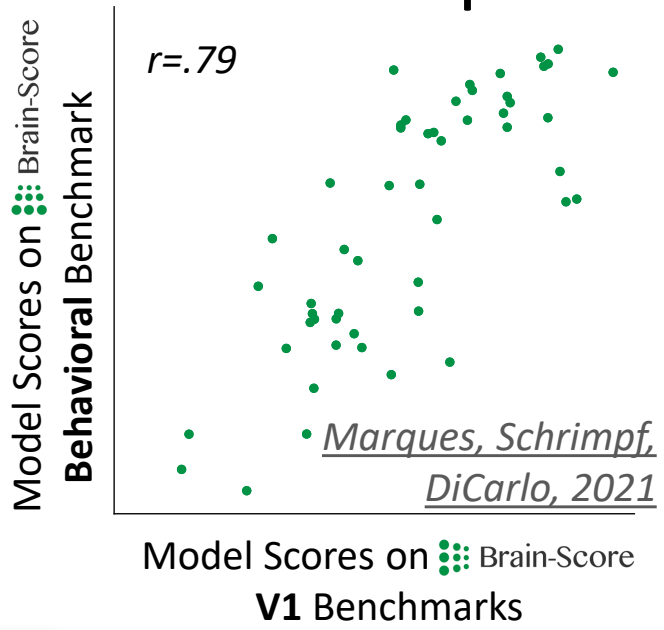
1. **Building ever more accurate unified ventral stream models**

2. **Relate brain benchmarks to one another.**
Improve V1 component of model
→ improve behavior

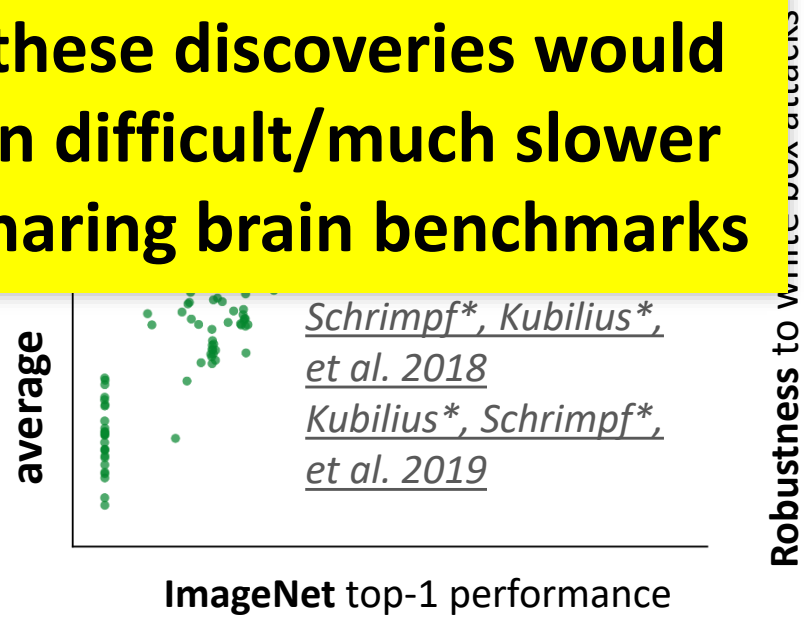
3. **Relate brain benchm… to engineering desiderata**
ImageNet → Brain-Score;

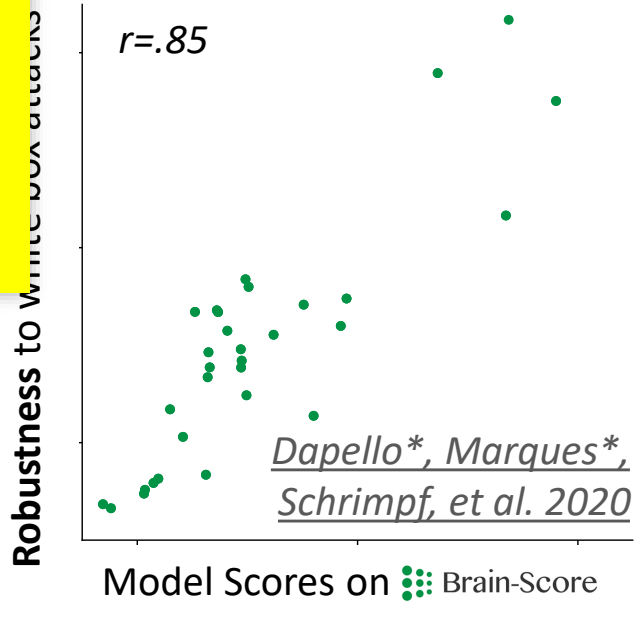
More V1-like
→ improved robustness





Many of these discoveries would have been difficult/much slower without sharing brain benchmarks

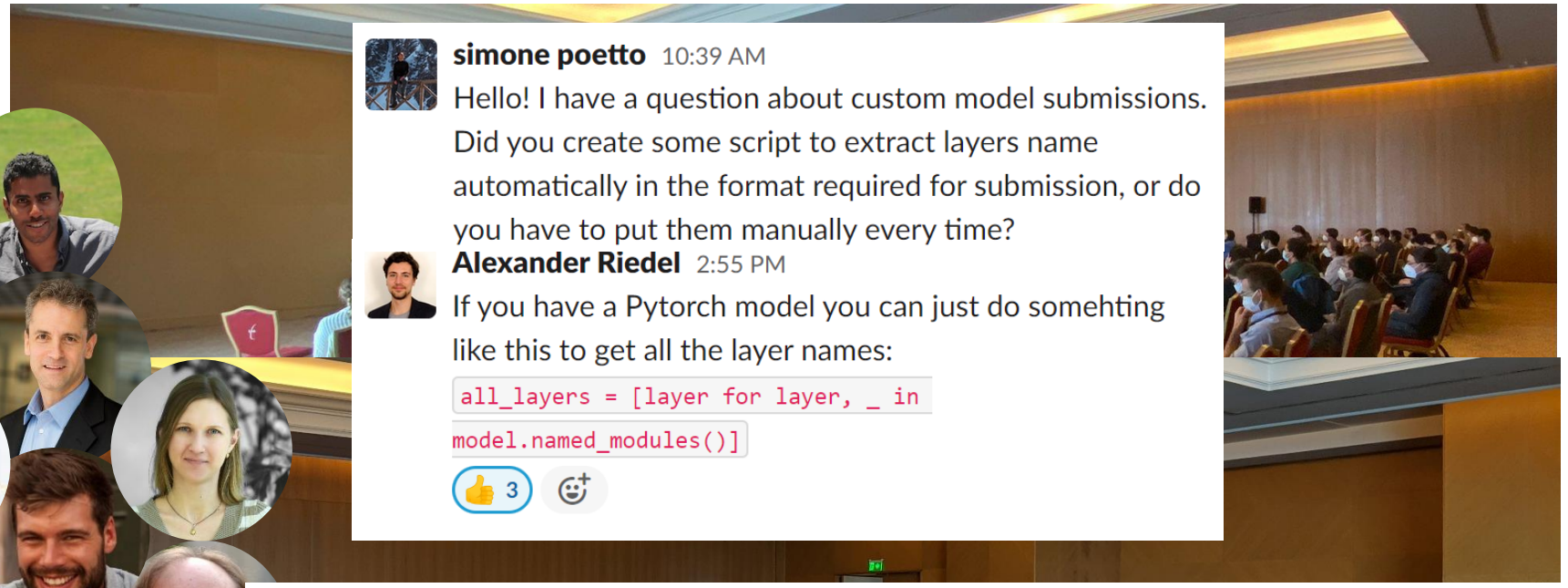




The Brain-Score community is essential

- Assembling data benchmarks to sufficiently evaluate brain models is insurmountable for any single lab
- Building better and better brain models under many axes (e.g. different neural regions, different behaviors) is insurmountable for any single lab
- Community involvement will scale much better than any single effort (hence the swag)

The Brain-Score community is growing



simone poetto 10:39 AM
Hello! I have a question about custom model submissions. Did you create some script to extract layers name automatically in the format required for submission, or do you have to put them manually every time?

Alexander Riedel 2:55 PM
If you have a Pytorch model you can just do somehting like this to get all the layer names:

```
all_layers = [layer for layer, _ in model.named_modules()]
```

👍 3

Aakashagr commented on Dec 28, 2021 — Member

Hello,

I receive the following error while installing the dependencies

ERROR: Could not find a version that satisfies the requirement tensorflow==1.15 (from model-tools) (from versions: 2.5.0rc0, 2.5.0rc1, 2.5.0rc2, 2.5.0rc3, 2.5.0, 2.5.1, 2.5.2, 2.6.0rc0, 2.6.0rc1, 2.6.0rc2, 2.6.0, 2.6.1, 2.6.2, 2.7.0rc0, 2.7.0rc1, 2.7.0, 2.8.0rc0)
ERROR: No matching distribution found for tensorflow==1.15

raj-magesh commented yesterday

Hey Aakash, in case you didn't solve this, try using Python 3.7.

👍 1

Community problem-solving at this workshop

Architecture

Kanaka Rajan: brain-wide region interactions	Uygar Sümbül: cell types	Panayiota Poirazi: dendrites

Learning rules

Kim Stachenfeld & Gabriela Michel: continual learning	Aaron Batista: motor learning	Andrew Saxe: multi-task

Adult inference

Kohitij Kar: cross-animal ventral similarity	Gabriel Kreiman: visual search	Arash Afraz: perceptography
Niko Kriegeskorte: adjudicating models	Carlos Ponce: dissecting representations	Jennifer Groh: multi-stimulus encoding
Hendrikje Nienborg: non-retinal input	Ramanujan Raghavan: V1 & V2 scene statistics	Eero Simoncelli: modeling V1 & V2

Community problem-solving at this workshop

Architecture

Kanaka Rajan: brain-wide region interactions

Uygar Sümbül: cell types

Panayiota Poirazi: dendrites

Learning rules

Kim Stachenfeld: Continual Michel: continual learning

Andrew Saxe: multi-task

Adult inference

Kohitij Kar: cross-animal ventral similarity

Gabriel Kreiman: visual search

Arash Afraz: perceptography

Niko Kriegeskorte: adjudicating models

Jennifer Groh: multi-stimulus encoding

Hendrikje Nienborg: non-retinal input

Ramanujan Raghavan: V1 & V2 scene statistics

Eero Simoncelli: modeling V1 & V2

Common themes

inter-animal ceiling

models informing experiments -- and vice-versa

where are current models falling short?

how can we separate models?

Lots of engagement from modelers

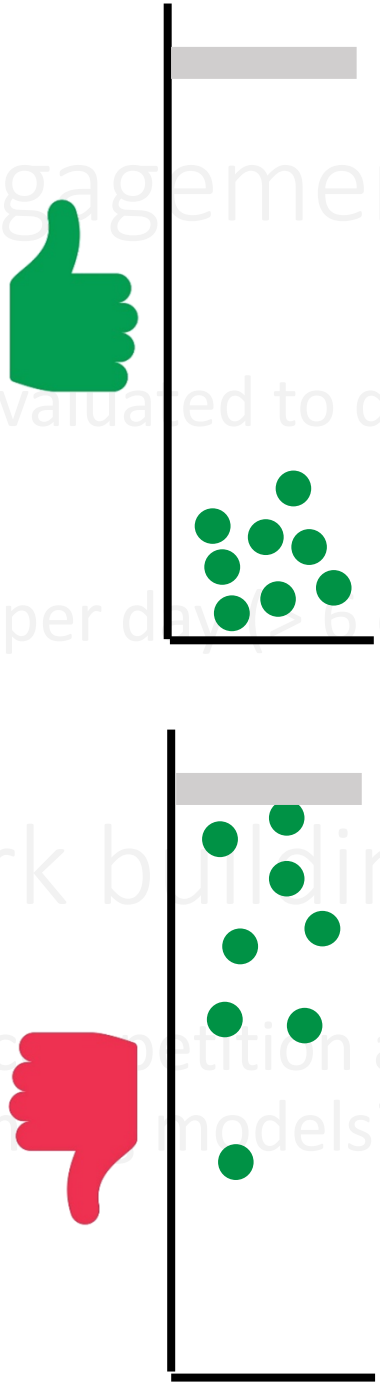
> 600 models evaluated to date

> 1 submission per day (> 6 during competition)

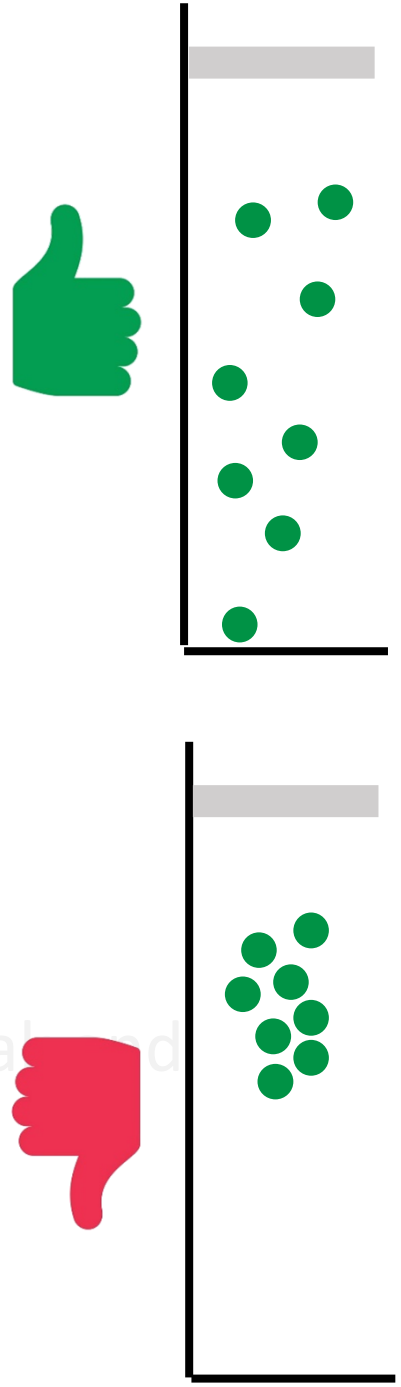
Benchmark building less active

Can we have a competition around benchmarks that break and differentiate among models?

“disruption score” 1: break the models



“disruption score” 2: separate the models



“The future of Brain-Score”

- Discovery of relationships across many different brain (and engineering) benchmarks
- More diverse neural and behavioral benchmarks
- This will only work as a community – need especially more model-breaking

What is the next step to make this cycle go faster?

Brain and Cognitive Sciences

Computational Models

What is the next step to make this cycle go faster?

Brain and Cognitive Sciences

Computational Models

How do we get the best benchmarks for model guidance and constraints?

1. **What is one key benchmark missing from Brain-Score right now?**
2. **What is preventing you from contributing it?**
3. **How to weight across benchmarks / decide on metrics and data?**