Deep generative networks as models of the visual system

Thomas Naselaris
Department of Neuroscience
Medical University of South Carolina (MUSC)
Charleston, SC
\[ t = 0 \]

While \( \text{dead} == \text{False} \):

\[
\text{thought}[t] = f(\text{thought}[0:t], \text{world}[0:t], \text{plans}[0:t])
\]

If \( \text{thought}[t] \) is fatal:

\[
\text{dead} = \text{True}
\]

Else:

\[
t += 1
\]

Infer the human algorithm.
What should the human (visual) algorithm do?

Arbitrary queries over representations
Does the dog have pointy ears?

A dog is there.

"clamped"
vision

mental imagery
Breedlove, St-Yves, Naselaris et al., *in rev.*
HOW TO TEST NETWORK AGAINST HUMAN BRAINS?
An experiment:

Cue: “ababie”

Picture:

Breedlove, St-Yves, Naselaris et al., in rev.
Breedlove, St-Yves, Naselaris et al., *in rev.*
Breedlove, St-Yves, Naselaris et al., in rev.
Imagine objects
Breedlove, St-Yves, Naselaris et al., *in rev.*
Breedlove, St-Yves, Naselaris et al., in rev.
Breedlove, St-Yves, Naselaris et al., *in rev.*
Prediction accuracy maps for visual and imagery encoding models

**Visual** encoding model (vEM) predicting voxel-wise brain activity during **visual** task

**Imagery** encoding model (iEM) predicting voxel-wise brain activity during **imagery** task
Tuning to **seen** and **imagined** spatial frequencies

Breedlove, St-Yves, Naselaris et al., *in rev.*
Receptive fields for **seen** and **imagined** stimuli

Breedlove, St-Yves, Naselaris et al., *in rev.*
Receptive fields for seen and imagined stimuli

RF size shift

RF eccentricity shift

Breedlove, St-Yves, Naselaris et al., in rev.
A deep generative model can predict differences in encoding of seen and mental images.
BUT IS THERE A DEEP GENERATIVE MODEL THAT CAN ACCURATELY PREDICT ACTIVITY DURING VISION OF NATURAL SCENES?
A DCNN-based encoding model yields more accurate predictions of brain activity in all visual areas than an encoding model based on a state-of-the-art deep generative network.
SO IS THAT A “NO” ON THE GENERATIVE MODEL IDEA?
PERHAPS THE “RIGHT” GENERATIVE MODEL IS HARD TO LEARN FROM IMAGE DATA ALONE.
MIGHT WE INFER IT DIRECTLY FROM BRAIN RESPONSES?
Natural Scenes Dataset
IT’S NOT YET CLEAR IF THIS WILL WORK. BUT IT’S CLEAR THAT MORE DATA REALLY HELPS
DCNN- vs. Gabor-based encoding models, ~1.5K data samples from vim-1

DCNN- vs. Gabor-based encoding models, ~5K data samples from the (incomplete) NSD
DCNN- vs. Gabor-based encoding models, ~5K data samples from the (incomplete) NSD

Data-driven vs. DCNN-based encoding models, ~5K data samples from the (incomplete) NSD
TAKE-HOME

THE VISUAL SYSTEM CAN POSE AND ANSWER MANY DIFFERENT QUERIES. SO SHOULD OUR MODELS.

A DEEP GENERATIVE MODEL CAN PREDICT DIFFERENCES IN ENCODING OF SEEN AND MENTAL IMAGES…
TAKE-HOME

...BUT CANNOT PREDICT RESPONSES TO NATURAL SCENES AS ACCURATELY AS MODELS BASED ON A DISCRIMINATIVE NETWORK.

WE NEED BETTER THEORY. AND MORE DATA.

MORE DATA IS ON THE WAY.
Acknowledgments

MUSC Naselaris Lab

Zahraa Sabra
Jesse Breedlove
Maggie Mae Mell
Ghislain St-Yves

CMRR

Cheryl Olman
Kendrick Kay

Funding
NIH R01 EY023384
BRAIN N00531701
NSF IIS-1822683
NSD Collaborators

- **Kendrick Kay, Assistant Professor, CMRR**
- **Thomas Naselaris, Assistant Professor, Medical University of South Carolina**
- **Emily Allen, Postdoctoral Associate, CMRR**
- **Yihan Wu, Graduate Student of Wilma Koutstaal, CCS**
- **Ben Hutchinson, Assistant Professor, University of Oregon**
- **Ariel Rokem, Senior Data Scientist, University of Washington**
2019 Conference on Cognitive Computational Neuroscience

13-16 September 2019  Berlin, Germany

https://ccneuro.org/2019/