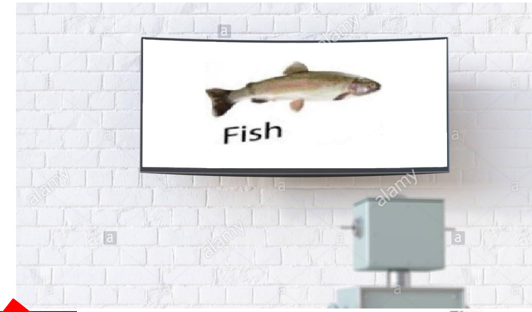
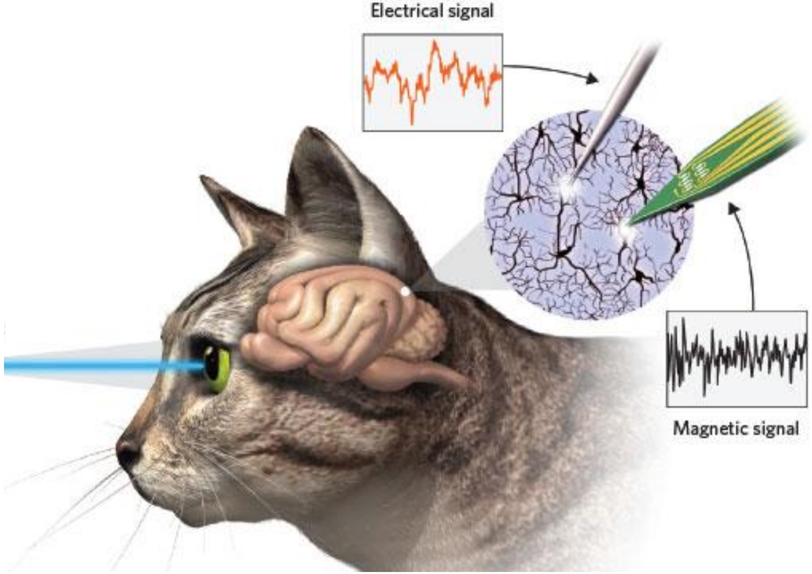
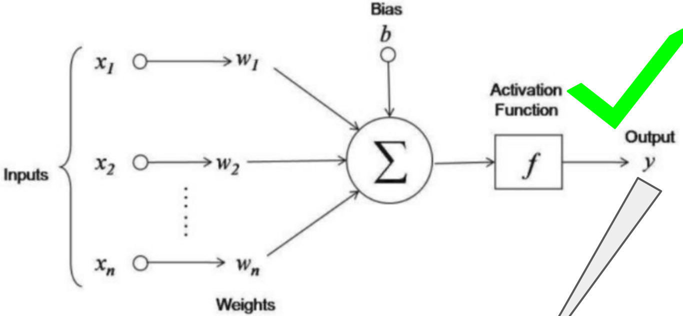
Comparing activations in biological and artificial neural networks

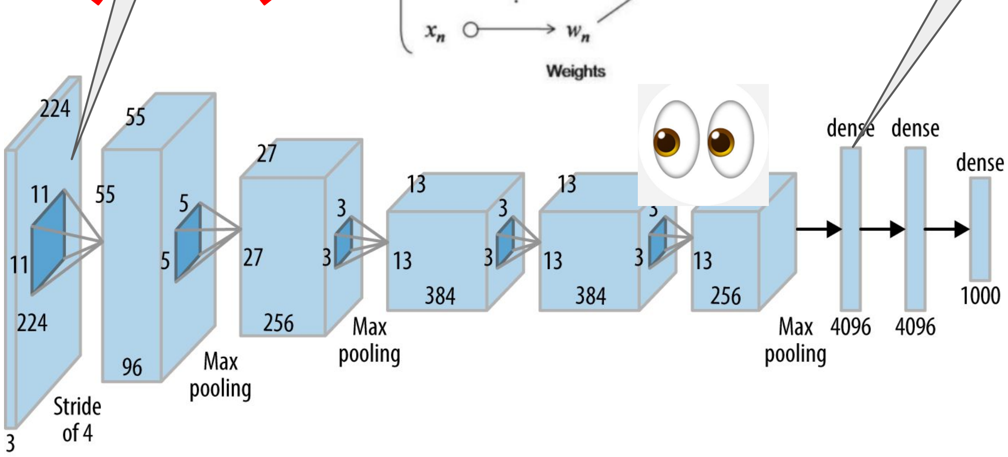
Jessica AF Thompson
MAIN DL Training
November 15, 2019
j.thompson@umontreal.ca

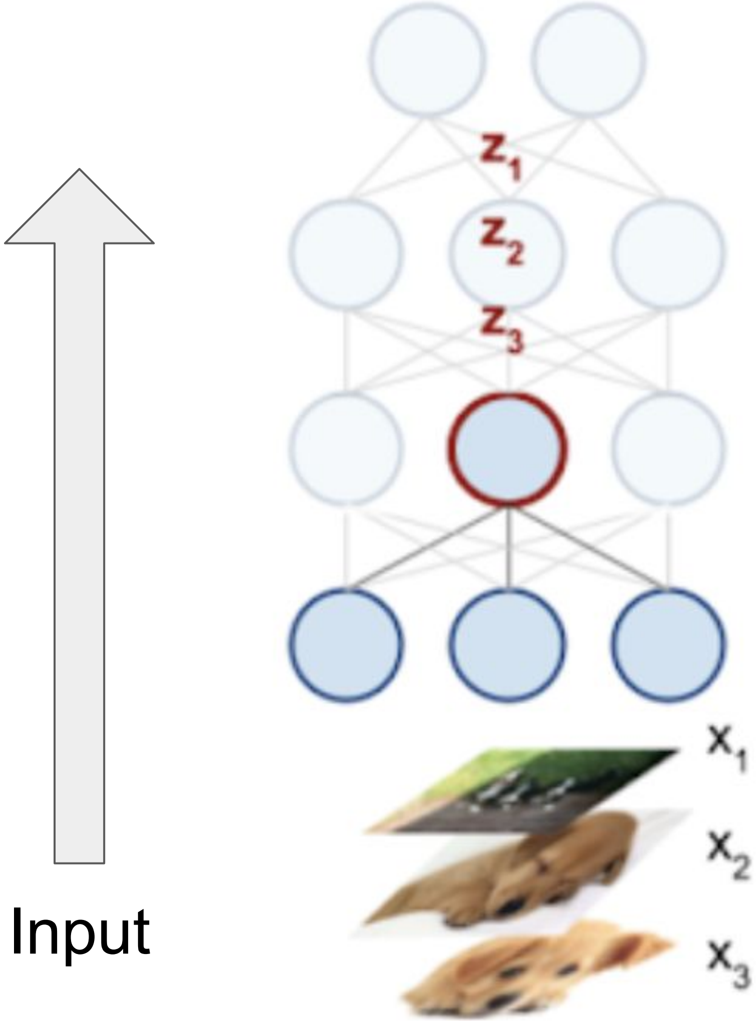

Fish

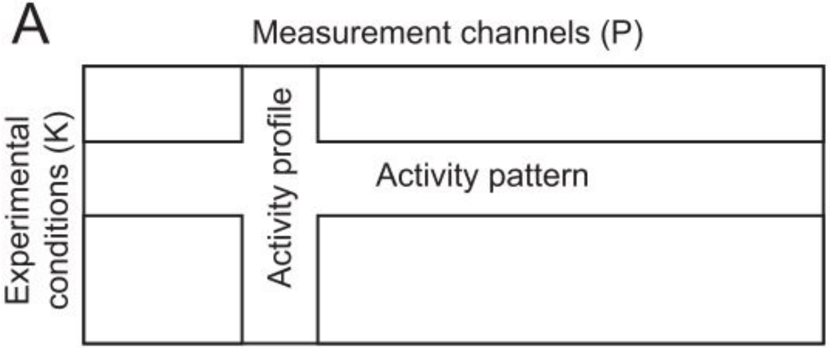

Fish


Electrical signal
Magnetic signal


Bias
b
x1
w1
x2
w2
xn
wn
Inputs
Weights
Σ
Activation Function
f
Output
y

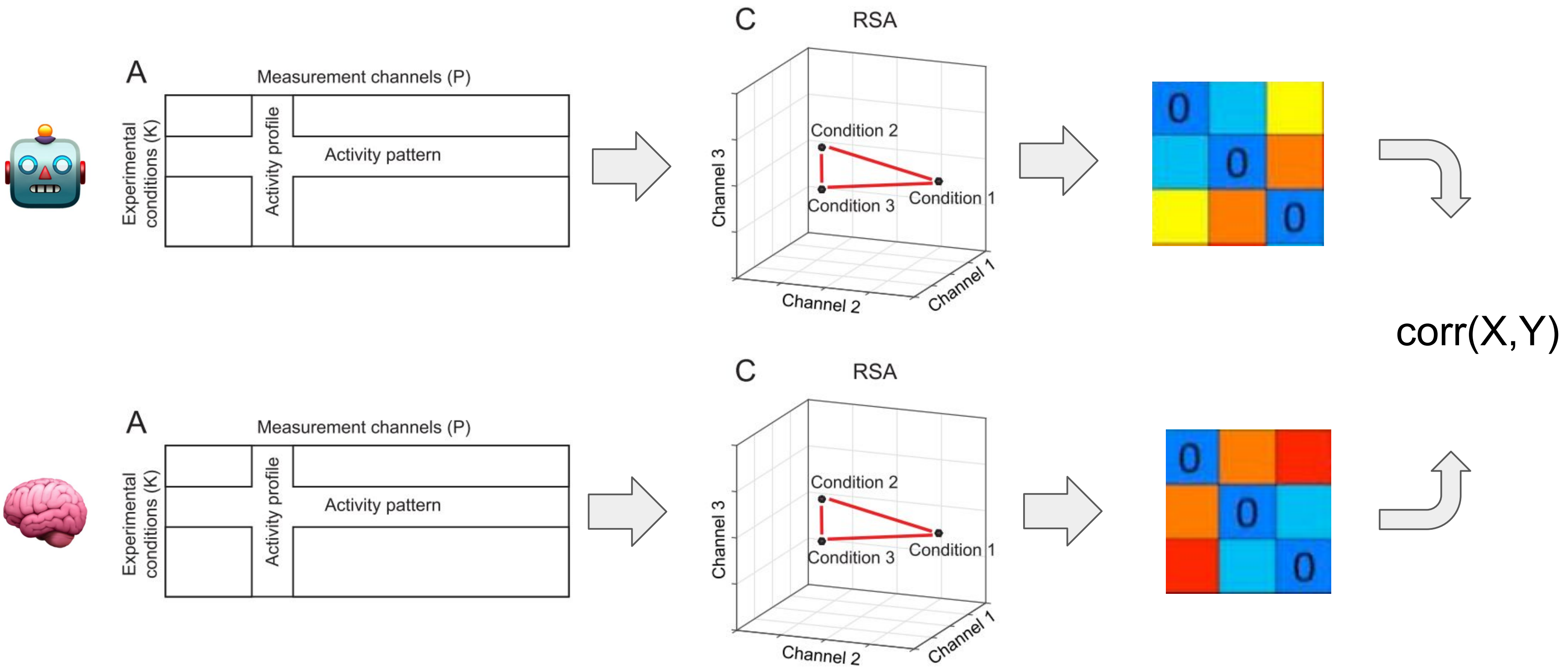

224
224
3
11
11
Stride of 4
55
55
96
5
5
Max pooling
27
27
256
3
3
Max pooling
13
13
384
3
3
13
13
384
3
13
256
Max pooling
dense
dense
dense
4096
4096
1000





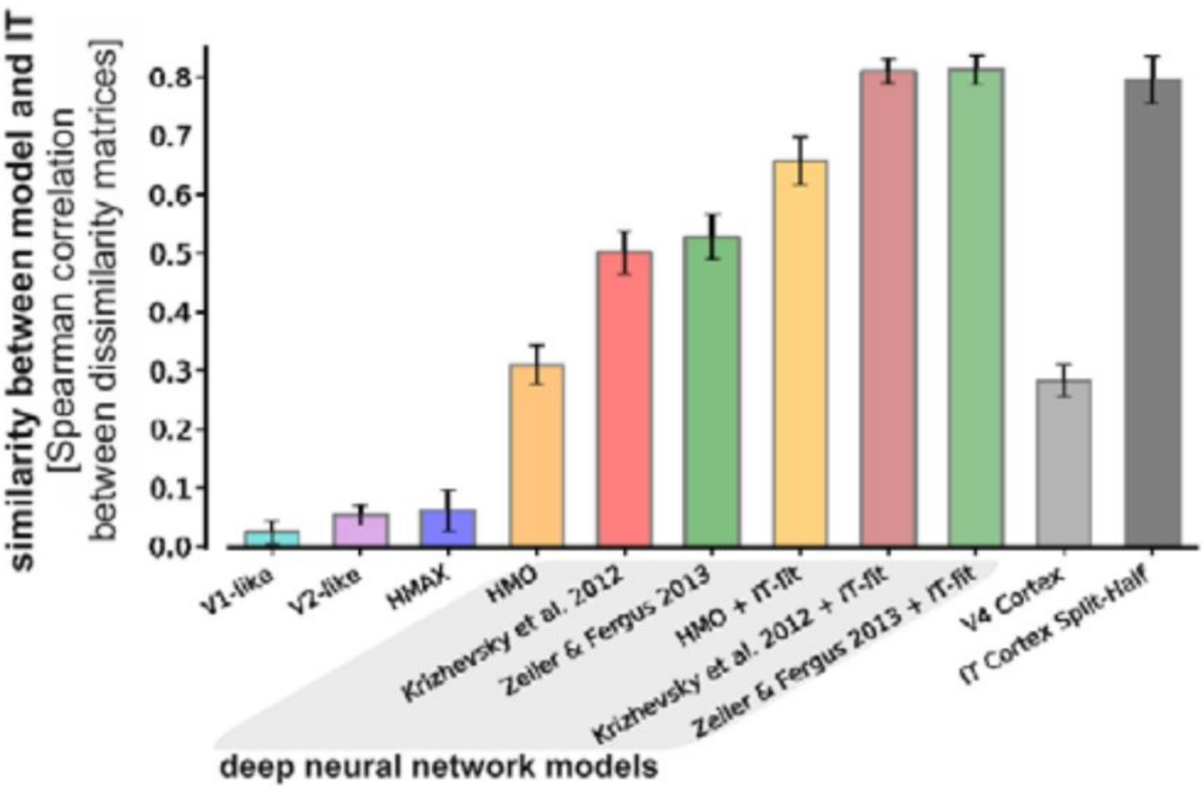
Diedrichsen, J., & Kriegeskorte, N. (2017)

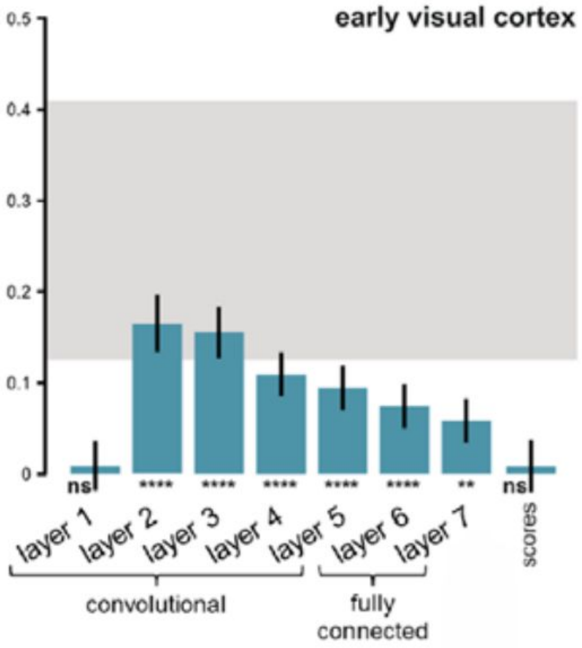
Representational Similarity Analysis (RSA)



Diedrichsen, J., & Kriegeskorte, N. (2017); Kriegeskorte, N., Mur, M., & Bandettini, P. (2008).

RSA to compare DNNs to the visual pathway





Kriegeskorte, N.(2015).

Responses

“Deep neural networks are uninterpretable and therefore can’t help us understand the brain”

“"What I cannot create, I do not understand" -Richard Feynman”

“Machine learning models have nothing to do with the brain”

“Convolutional neural networks were inspired by the mammalian visual system”

“You’re just replacing one black box with another”

“Could a neuroscientist understand a microprocessor?”

Questions

1. What do we learn from comparing artificial and biological neural networks?
    a. What kinds of questions does this analysis answer?
    b. How does this type of analysis compare to existing analyses approaches?
    c. Does it provide a new way of answering existing questions or does it ask new questions?
2. How does this type of science progress?
    a. How do we get closer to truth?
    b. What do we want the product of our science to be? What is success?
3. What is the role of the artificial neural network in this framework?
    a. Is it an analysis tool, a computational model, or a model organism?
4. Is this approach better than other approaches?

Outline

- Situate
  - How does this approach fit into the landscape of other approaches?
- Literature review
  - Setting the scene
  - Deep networks are good models of the brain
  - Thoughts and feelings
  - Now
- Questions recap and conclusions

Topics at the intersection of AI and neuroscience

Areas of study

1. Representations
   a. How is relevant information encoded?
   b. How is information being transformed?
2. Architectures
   a. How are different components put together?
3. Training algorithms
   a. Learning rules and optimization
   b. Cost functions
   c. Curriculum

most neuroscience

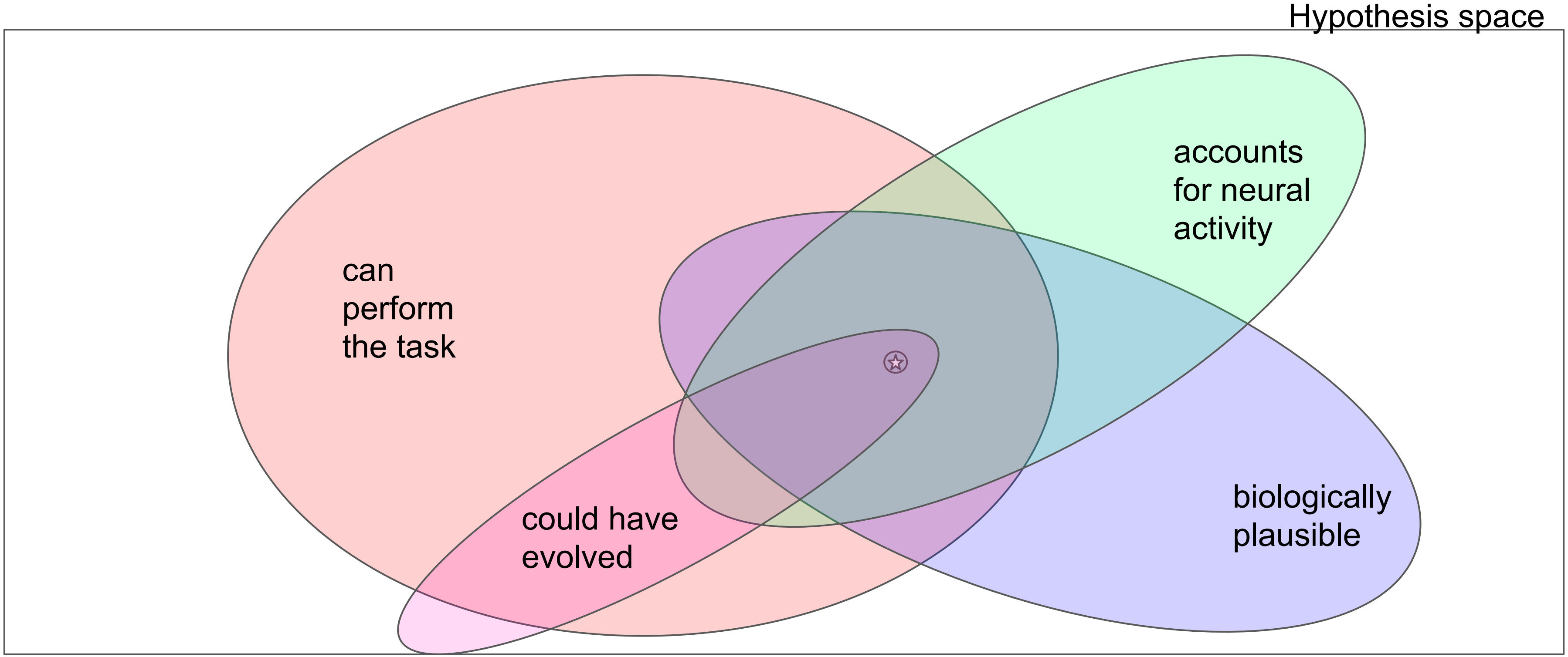
most machine learning

science of intelligence

Two Scientific Approaches

- Null hypothesis significance testing
  - Searching for reliable effects
  - e.g. classical fMRI GLM analysis
- Model comparison
  - Adjudicate among competing candidate models of some process/phenomenon
  - Computational neuroscience
  - e.g. Neural encoding analysis, often
  - e.g. Comparing artificial and biological network activations, usually

Model Comparison Approach



RESEARCH ARTICLE

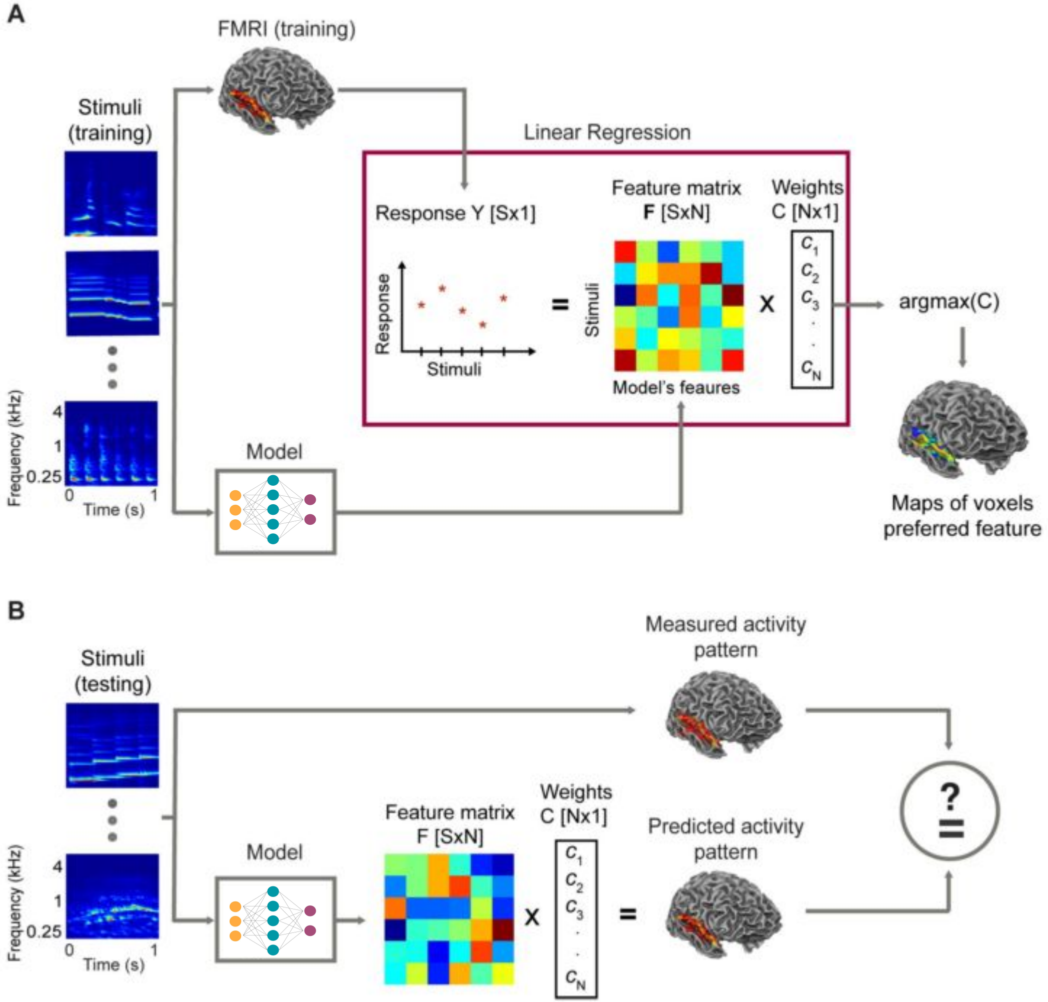
Scientific discovery in a model-centric framework: Reproducibility, innovation, and epistemic diversity

Berna Devezer[1,5]*, Luis G. Nardin[2,5], Bert Baumgaertner[3,5], Erkan Ozge Buzbas[4,5]

- Simulation of scientific discovery in a model-centric approach
    - Innovative research speeds up the discovery of scientific truth by facilitating the exploration of model space
    - Epistemic diversity optimizes across desirable properties of scientific discovery

Model Comparison in Functional Neuroscience

- Encoding analysis
  - Hypotheses about the nature of neural representations (i.e. neural code)
- Comparison with DNN activity
  - Hypotheses about what architectures and training procedures lead to brain-like representations



Santoro, R. (2014). The Computational Architecture of the Human Auditory Cortex.

Statistical tools to compare two sets of variables

- Linear Regression
- Representational Similarity Analysis
- Pattern Component Modeling
- Canonical Correlation Analysis
  - Singular Vector CCA
  - Projection Weighted CCA
- Centered Kernel Alignment
- Hyperalignment

Applications

- Questions about representations in artificial and biological neural networks
- Questions about architecture in artificial and biological neural networks
- Questions about learning in artificial and biological neural networks
- Comparing brains to models, comparing models to models, comparing brains to brains.

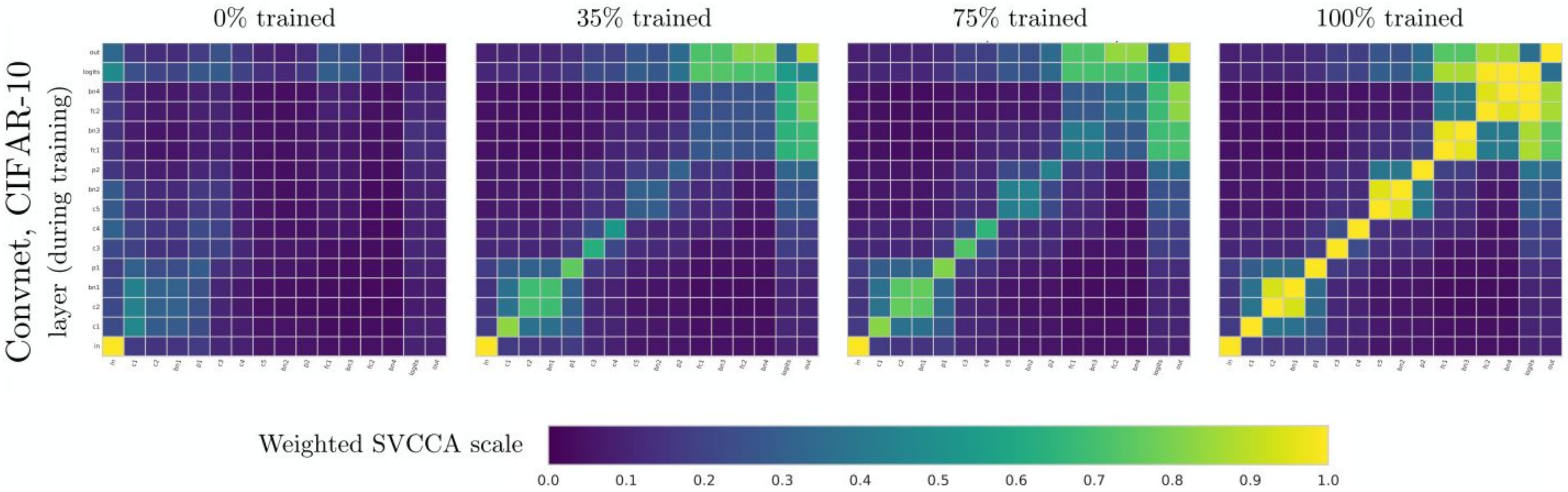
An overview of functional alignment in artificial and biological neural networks: Current recommendations and open questions

Elizabeth DuPre (elizabeth.dupre@mail.mcgill.ca)
Montreal Neurological Institute, McGill University
Montreal, QC, Canada

Jean-Baptiste Poline (jean-baptiste.poline@mcgill.ca)
Montreal Neurological Institute, McGill University
Montreal, QC, Canada

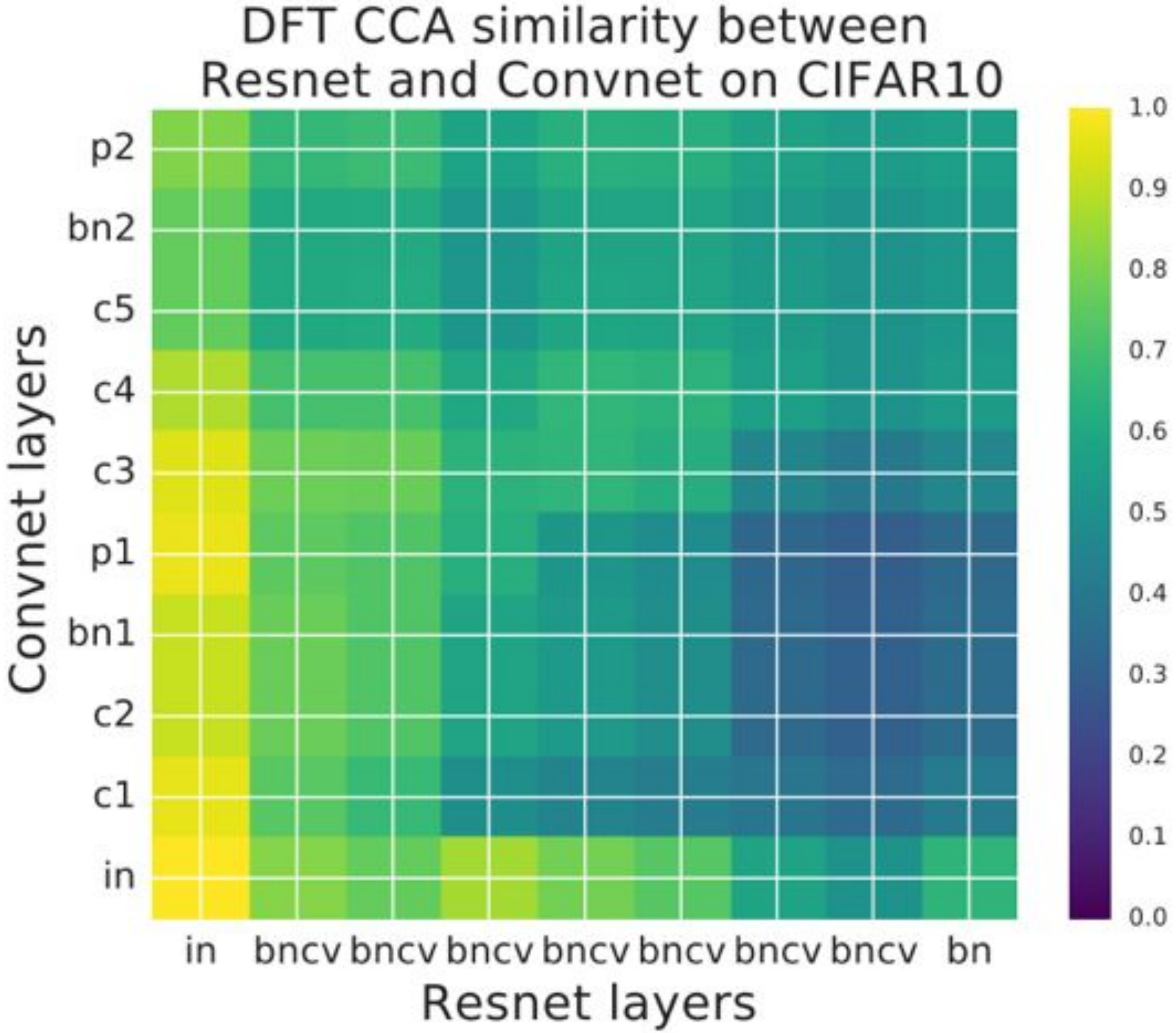
Check out her poster at MAIN!

Using SVCCA to study learning dynamics in deep networks



Raghu, M., Gilmer, J., Yosinski, J., & Sohl-Dickstein, J. (2017). SVCCA: Singular Vector Canonical Correlation Analysis for Deep Understanding and Improvement. NeurIPS.

Comparing representations in two different architectures



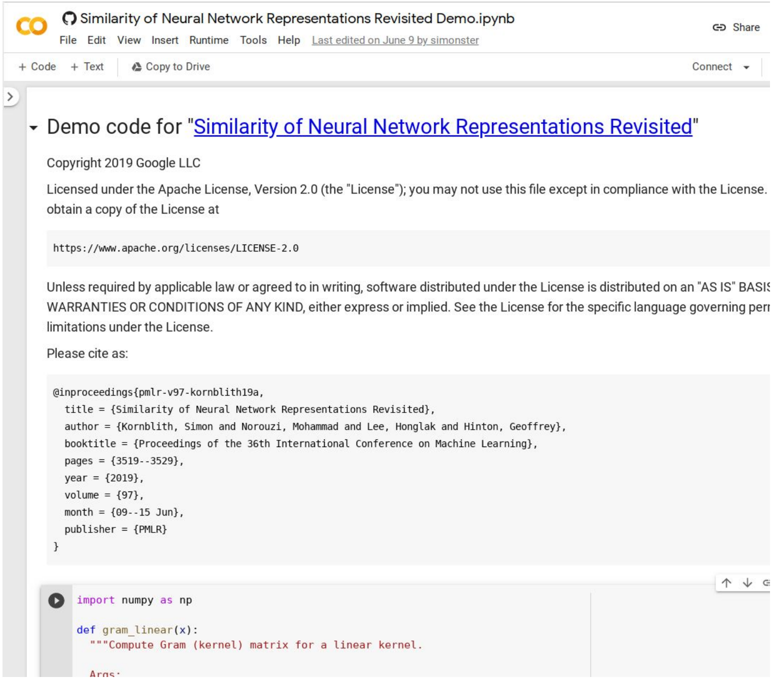
Morcos, A. S., Raghu, M., & Bengio, S. (2018). Insights on representational similarity in neural networks with canonical correlation. NeurIPS.

Similarity of Neural Network Representations Revisited

Simon Kornblith [1] Mohammad Norouzi [1] Honglak Lee [1 2] Geoffrey Hinton [1]

- [Paper]
- [Colab]

		Invariant to		
Similarity Index	**Formula**	**Invertible Linear Transform**	**Orthogonal Transform**	**Isotropic Scaling**
Linear Regression (R^2_{LR})	$\lVert Q_Y^{\mathrm{T}} X\rVert_{\mathrm{F}}^2/\lVert X\rVert_{\mathrm{F}}^2$	Y Only	✓	✓
CCA (R^2_{CCA})	$\lVert Q_Y^{\mathrm{T}} Q_X\rVert_{\mathrm{F}}^2/p_1$	✓	✓	✓
CCA ($\bar{\rho}_{\mathrm{CCA}}$)	$\lVert Q_Y^{\mathrm{T}} Q_X\rVert_*/p_1$	✓	✓	✓
SVCCA (R^2_{SVCCA})	$\lVert (U_Y T_Y)^{\mathrm{T}} U_X T_X\rVert_{\mathrm{F}}^2/\min(\lVert T_X\rVert_{\mathrm{F}}^2, \lVert T_Y\rVert_{\mathrm{F}}^2)$	In a Subspace	✓	✓
SVCCA ($\bar{\rho}_{\mathrm{SVCCA}}$)	$\lVert (U_Y T_Y)^{\mathrm{T}} U_X T_X\rVert_*/\min(\lVert T_X\rVert_{\mathrm{F}}^2, \lVert T_Y\rVert_{\mathrm{F}}^2)$	In a Subspace	✓	✓
PWCCA	$\sum_{i=1}^{p_1} \alpha_i \rho_i/\lVert\alpha\rVert_1,\ \alpha_i = \sum_j \lvert\langle \mathbf{h}_i, \mathbf{x}_j\rangle\rvert$	✗	✗	✓
Linear HSIC	$\lVert Y^{\mathrm{T}} X\rVert_{\mathrm{F}}^2$	✗	✓	✗
Linear CKA	$\lVert Y^{\mathrm{T}} X\rVert_{\mathrm{F}}^2/(\lVert X^{\mathrm{T}} X\rVert_{\mathrm{F}}\lVert Y^{\mathrm{T}} Y\rVert_{\mathrm{F}})$	✗	✓	✓
RBF CKA	$\mathrm{tr}(KHLH)/\sqrt{\mathrm{tr}(KHKH)\mathrm{tr}(LHLH)}$	✗	✓	✓*



2007-2012

Talking about neural processes with the same language used to talk about DNNs

Opinion TRENDS in Cognitive Sciences Vol.11 No.8

ELSEVIER

Full text provided by www.sciencedirect.com
ScienceDirect

Untangling invariant object recognition

James J. DiCarlo and David D. Cox

McGovern Institute for Brain Research, and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Despite tremendous variation in the appearance of visual objects, primates can recognize a multitude of objects, each in a fraction of a second, with no apparent effort. However, the brain mechanisms that enable this fundamental ability are not understood. Drawing on ideas from neurophysiology and computation, we present a graphical perspective on the key computational challenges of object recognition, and argue that the format of neuronal population representation and a property that we term 'object tangling' are central. We use this perspective to show that the primate ventral visual processing stream achieves a particularly effective solution in which single-neuron invariance is not the goal. Finally, we speculate on the key neuronal mechanisms that could enable this solution, which, if understood, would have far-reaching implications for cognitive neuroscience.

the table, bring fort
brain, and pull th
framework. Below,
vide intuition abou
that the primate
duces a particular
poral (IT) cortex, a
stream approaches
that some approa
distract from, unde

What is object rec
We define object r
discriminate each
objects ('categoriza
materials, textures

Neuron
Perspective

How Does the Brain Solve Visual Object Recognition?

James J. DiCarlo,[1,*] Davide Zoccolan,[2] and Nicole C. Rust[3]
[1]Department of Brain and Cognitive Sciences and McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2]Cognitive Neuroscience and Neurobiology Sectors, International School for Advanced Studies (SISSA), Trieste, 34136, Italy
[3]Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA
*Correspondence: dicarlo@mit.edu
DOI 10.1016/j.neuron.2012.01.010

Mounting evidence suggests that 'core object recognition,' the ability to rapidly recognize objects despite substantial appearance variation, is solved in the brain via a cascade of reflexive, largely feedforward computations that culminate in a powerful neuronal representation in the inferior temporal cortex. However, the algorithm that produces this solution remains poorly understood. Here we review evidence ranging from individual neurons and neuronal populations to behavior and computational models. We propose that understanding this algorithm will require using neuronal and psychophysical data to sift through many computational models, each based on building blocks of small, canonical subnetworks with a common functional goal.

2014-2016

DNNs are good models of the primate visual (and maybe auditory) sensory systems

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation

Seyed-Mahdi Khaligh-Razavi*, Nikolaus Kriegeskorte*

Medical Research Council, Cognition and Brain Sciences Unit, Cambridge, United Kingdom

Abstract

Inferior temporal (IT)
vision models, althou
internal representatio
computational model
the IT representation
VisNet) along with se
neural network). We
RDMs obtained from
stimuli (not used in t
clustering of represer
in terms of their with
between IT and man
unsupervised models
labeled images, reach
the IT data. Combinin
the margin between a
explained our IT dat
supervised learning t

Behavioral/Cognitive

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

Umut Güçlü and Marcel A. J. van Gerven

Radboud University, Donders Institute for Brain, Cogni

Converging evidence suggests that the primate v
areas. We quantitatively show that there indeed
brain. This was achieved by mapping thousands c
network. Our approach also revealed a fine-grain
allowed decoding of representations from hum
developed approach. Stimulus features that suc
plicitly tuned for object categorization. This prov
the functional organization of the primate ventra

Key words: deep learning; functional magnetic r

Using goal-driven deep learning models to understand sensory cortex

Daniel L K Yamins[1,2] & James J DiCarlo[1,2]

**Fueled by innovation in the computer vision and art
intelligence communities, recent developments in
computational neuroscience have used goal-driven
convolutional neural networks (HCNNs) to make str
modeling neural single-unit and population respons
visual cortical areas. In this Perspective, we review
progress in a broader modeling context and describ
the key technical innovations that have supported i
outline how the goal-driven HCNN approach can be
delve even more deeply into understanding the dev
and organization of sensory cortical processing.**

Brains on Beats

Umut Güçlü
Radboud University, Donders Institute for
Brain, Cognition and Behaviour
Nijmegen, the Netherlands
`u.guclu@donders.ru.nl`

Jordy Thielen
Radboud University, Donders Institute for
Brain, Cognition and Behaviour
Nijmegen, the Netherlands
`j.thielen@psych.ru.nl`

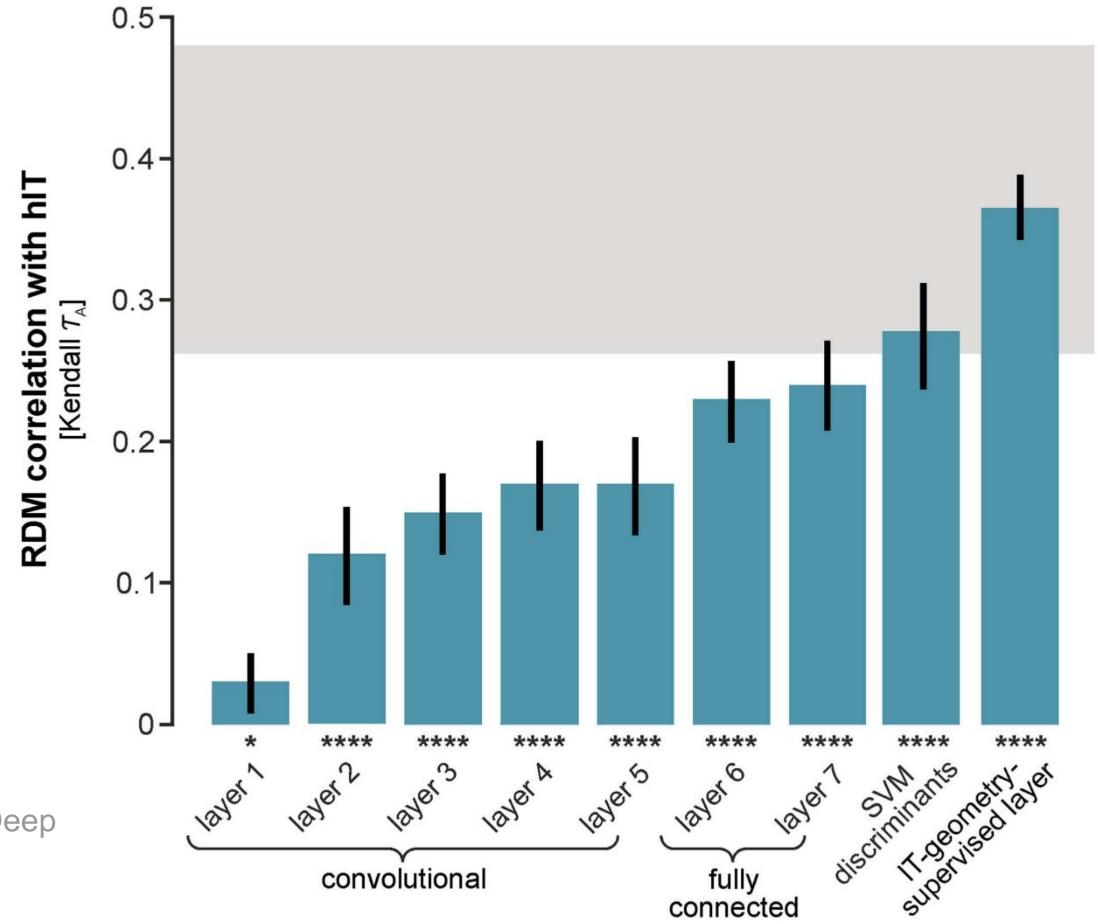
Michael Hanke*
Otto-von-Guericke University Magdeburg
Center for Behavioral Brain Sciences
Magdeburg, Germany
`michael.hanke@ovgu.de`

Marcel A. J. van Gerven†
Radboud University, Donders Institute for
Brain, Cognition and Behaviour
Nijmegen, the Netherlands
`m.vangerven@donders.ru.nl`

Abstract

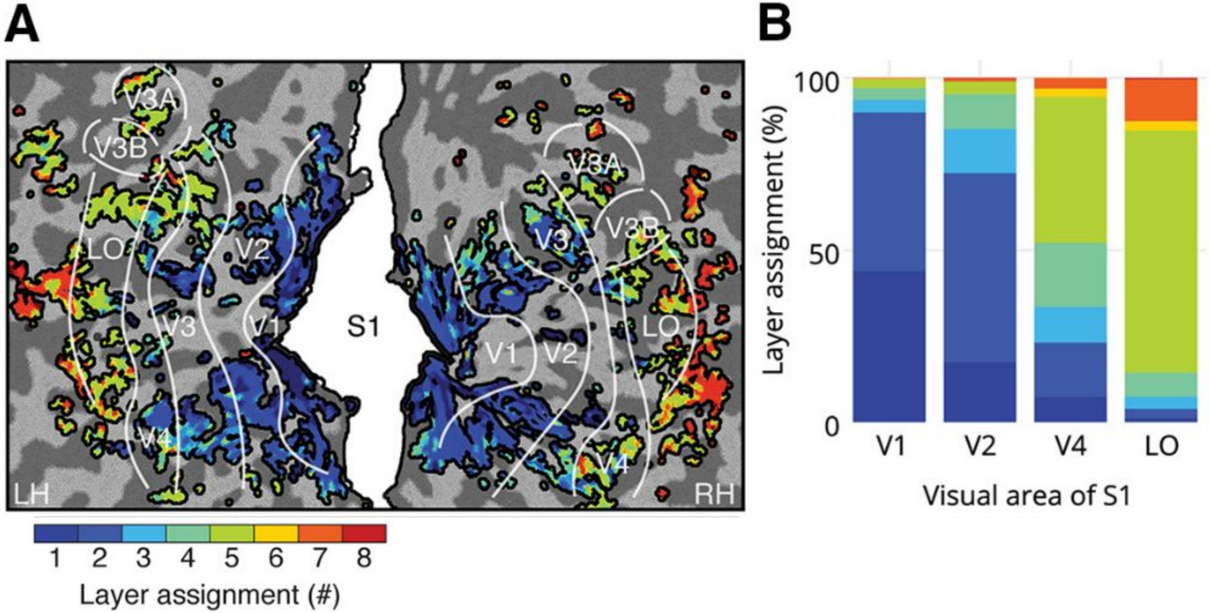
We developed task-optimized deep neural networks (DNNs) that achieved state-of-

“AlexNet, with features remixed and reweighted, fully explains data from human IT”



Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. PLoS Computational Biology, 10(11), e1003915.

“Layer assignments increase as a function of position on the occipital cortex”



Güçlü, U., & van Gerven, M. A. J. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. The Journal of Neuroscience

2016-2018

Thoughts and feelings

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Current Opinion in Neurobiology

Analyzing biological and artificial neural networks: challenges with opportunities for synergy?

David GT Barrett[1,3], Ari S Morcos[1,3,4] and Jakob H Macke[2]

Check for updates

frontiers in Computational Neuroscience

HYPOTHESIS AND THEORY
published: 14 September 2016
doi: 10.3389/fncom.2016.00094

Check for updates

How can deep learning advance computational modeling of sensory information processing?

Jessica A.F. Thompson[1,2], Yoshua Bengio[2], Elia Formisano[3], and Marc Schönwiesner[1,4]

[1] International Laboratory for Brain, Music and Sound, University of Montreal, Canada
[2] Montreal Institute for Learning Algorithms, Montreal, Canada
[3] Department of Cognitive Neuroscience, Maastricht University, Maastricht, Netherlands
[4] Institute for Biology, Leipzig University, Leipzig, Germany

Toward an Integration of Deep Learning and Neuroscience

Adam H. Marblestone[1], Greg Wayne[2] and Konrad P. Kording[3]*

[1] Synthetic Neurobiology Group, Massachusetts Institute of Technology, Media Lab, Cambridge, MA, USA, [2] Google Deepmind, London, UK, [3] Rehabilitation Institute of Chicago, Northwestern University, Chicago, IL, USA

2016-2018

Thoughts and feelings

Contents lists available at ScienceDirect
NeuroImage
journal homepage: www.elsevier.com/locate/neuroimage

Outlook on deep neural networks in computational cognitive neuroscience

Brandon M. Turner [a], Steven Miletić [b], Birte U. Forstmann [b,*]

[a] The Ohio State University, United States
[b] University of Amsterdam, The Netherlands

Contents lists available at ScienceDirect
NeuroImage
journal homepage: www.elsevier.com/locate/neuroimage

Fantastic DNimals and where to find them

H. Steven Scholte

Department of Brain & Cognition, University of Amsterdam, The Netherlands

Contents lists available at ScienceDirect
NeuroImage
journal homepage: www.elsevier.com/locate/neuroimage

A deeper understanding of the brain

Bryan Tripp

University of Waterloo, Canada

Contents lists available at ScienceDirect
NeuroImage
journal homepage: www.elsevier.com/locate/neuroimage

Review

Principles for models of neural information processing

Kendrick N. Kay

Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, Twin Cities, Minneapolis, MN, USA

Contents lists available at ScienceDirect
NeuroImage
journal homepage: www.elsevier.com/locate/neuroimage

Predict, then simplify

Jonas Kubilius [a,b,*]

[a] McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge 46-6161, USA
[b] Brain and Cognition, KU Leuven, Leuven, Belgium

Panel on Explaining Cognition, Brain Computation and Intelligent Behaviour

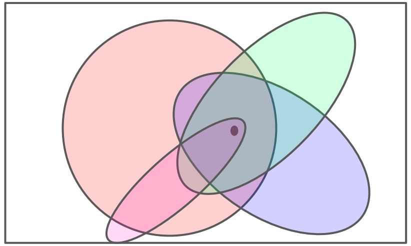
Question posed by Jim DiCarlo:

What is your definition of success?

Answers from Yann LeCun, Jackie Gottlieb, Josh Tenenbaum, and Nancy Kanwisher

Is it a problem?

- We need more clarity and consensus about the long term goals of our field.
  - What will be the form of adequate explanations of intelligent capacities?
- It’s not at all problematic that we have varied short-term goals. In fact it is probably beneficial!
- Good predictions of brain activity is not a sufficient condition for evaluating models. It is just one of several constraints on model space.



2019

- How good are these models *really*?
- Adding biological realism

How well do deep neural networks trained on object recognition characterize the mouse visual system?

> "no match between the hierarchy of mouse visual cortical areas and the layers of CNNs trained on object categorization."

> "Although [the network] achieves state-of-the-art performance, it is m

[6] Institute Bioinformatics and Medical I
† Autho
‡ Present address: Department of Co
* santiago.c

2019

- **How good are these models *really*?**
- Adding biological realism

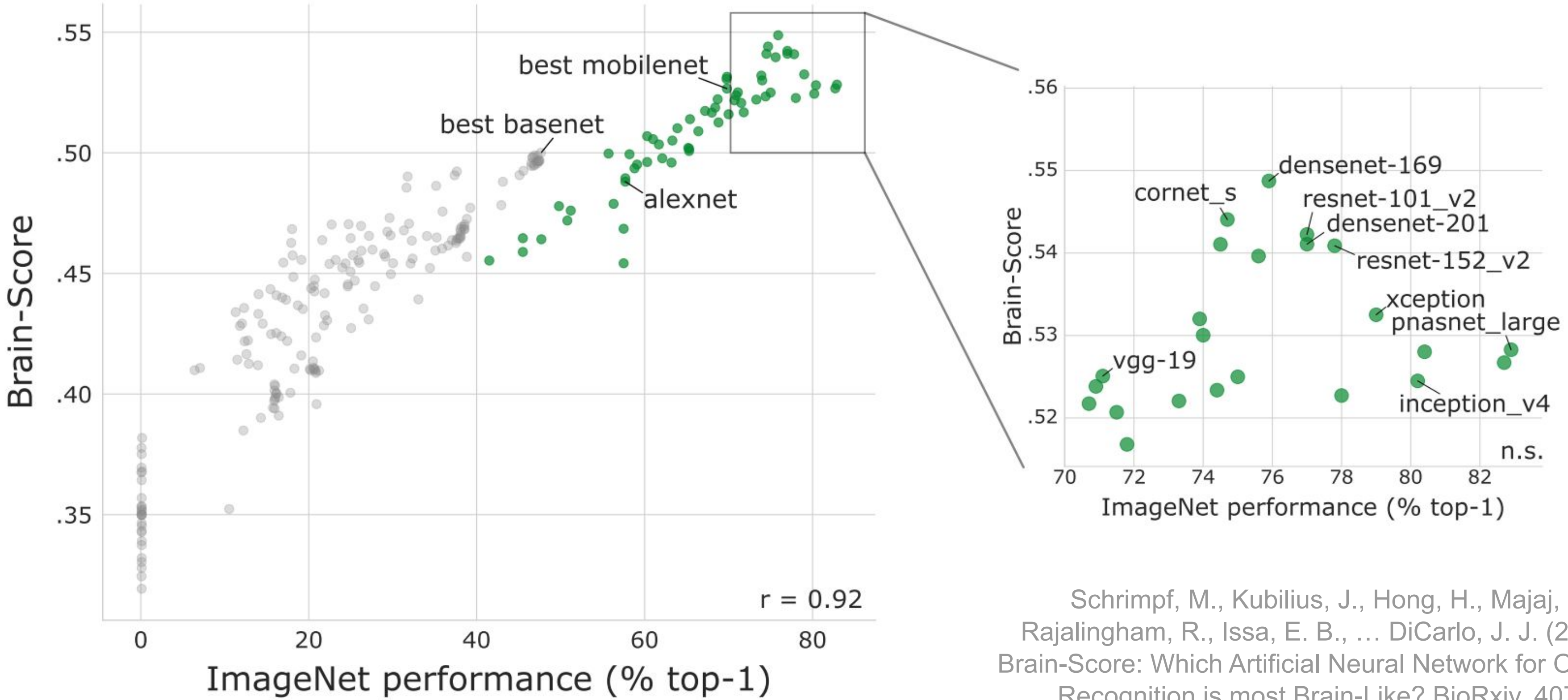
Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like?

Martin Schrimpf[*,1,2], Jonas Kubilius[*,3,4], Ha Hong[5], Najib J. Majaj[6], Rishi Rajalingham[1], Elias B. Issa[7], Kohitij Kar[1,3], Pouya Bashivan[1,3], Jonathan Prescott-Roy[1], Kailyn Schmidt[1], Daniel L. K. Yamins[8,9], and James J. DiCarlo[1,2,3]

[1]Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139
[2]Center for Brains, Minds and Machines, MIT, Cambridge, MA 02139
[3]McGovern Institute for Brain Research, MIT, Cambridge, MA 02139
[4]Brain and Cognition, KU Leuven, Leuven, Belgium
[5]Bay Labs Inc., San Francisco, CA 94102
[6]Center for Neural Science, New York University, New York, NY 10003
[7]Department of Neuroscience, Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10027
[8]Department of Psychology, Stanford University, Stanford, CA 94305
[9]Department of Computer Science, Stanford University, Stanford, CA 94305

Brain-Score

As deep ANNs continue to evolve, are they becoming more or less brain-like?



Schrimpf, M., Kubilius, J., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., … DiCarlo, J. J. (2018). Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like? BioRxiv, 407007.

Brain-Score

Leaderboard | About | Compare | Participate

Rank ▼	Model	average	V1: FreemanZiemba2013.V1-pls	V2: FreemanZiemba2013.V2-pls	V4: Majaj2015.V4-pls	IT: Majaj2015.IT-pls	IT-temporal: Kar2019-ost	behavior: Rajalingham2018-i2n	ImageNet: Deng2009-top1
1	CORnet-S *Kubilius et al., 2018*	.402	.146	.208	.593	.533	.384	.545	.747
2	densenet-169 *Huang et al., 2016*	.365	.198	.288	.618	.542	X	.543	.759
2	resnet-101_v1 *He et al., 2015*	.365	.207	.274	.600	.545	X	.561	.764
4	resnet-50_v1 *He et al., 2015*	.364	.208	.279	.611	.558	X	.526	.752
	resnet-152_v1								

Are Topographic Deep Convolutional Neural Networks Better Models of the Ventral Visual Stream?

Kamila Maria Jozwik (kmjozwik@mit.edu)
University of Cambridge and McGovern Institute for Brain Research, Center for Brains, Minds and Machines at Massachusetts Institute of Technology, 43 Vassar St
Cambridge, MA 02139 United States

Hyodong Lee (hyo@mit.edu)
Department of Electrical Engineering and Computer Science at MIT

Nancy Kanwisher (ngk@mit.edu)
McGovern Institute for Brain Research and Department of Brain and Cognitive Sciences at MIT

James J. Di
McGovern Institute for Brain Research a

2019

- How good are these models *really*?
- **Adding biological realism**

Do Biologically-Realistic Recurrent Architectures Produce Biologically-Realistic Models?

Grace W. Lindsay (gracewlindsay@gmail.com)

"Here we show that it is possible to incorporate more biologically realistic details, in the form of recurrent connections, into a standard convolutional neural network... In doing so, we show that certain architectural features— such as only allowing excitatory cells to be output cells—help replicate findings from the data and lead to different types of image representations. The architectural features that provide these benefits do not, however, necessarily make the image representations in the model more similar to that of V4 data. Reconciling these differences will be important."

What’s next?

- Other modalities
  - Audition
  - Language
- Formalizing our shared definition of long-term success
- Evaluation metrics: What does it mean to be ‘brain-like’?
- Experimental design
  - Collecting large amounts of data from individual subjects

Questions

1. What do we learn from comparing artificial and biological neural networks?
   a. What kinds of questions does this analysis answer?
   b. How does this type of analysis compare to existing analyses approaches?
   c. Does it provide a new way of answering existing questions or does it ask new questions?
2. How does this type of science progress?
   a. How do we get closer to truth?
   b. What do we want the product of our science to be?
3. What is the role of the artificial neural network in this framework?
   a. Is it an analysis tool, a computational model, or a model organism?
4. Is this approach better than other approaches?

Conclusion

- Comparing activations in biological and artificial neural networks is a promising approach to study the architectures and processes that support brain-like representations and the nature of representations in intelligent systems
- But it’s not just about chasing high accuracies
  - Learning how to build a neural network won’t teach you how to use them to do science
  - Science is not an engineering problem, no matter how much we want it to be
  - The (long term) goal of science is to generate scientific explanations, which is not the same as statistically explaining the variance in our data
- Epistemic diversity optimizes scientific discovery

Thank you for your attention

Questions?