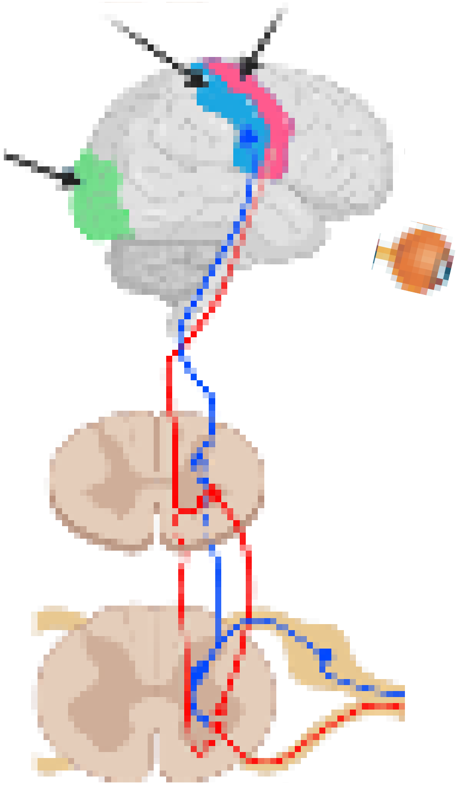
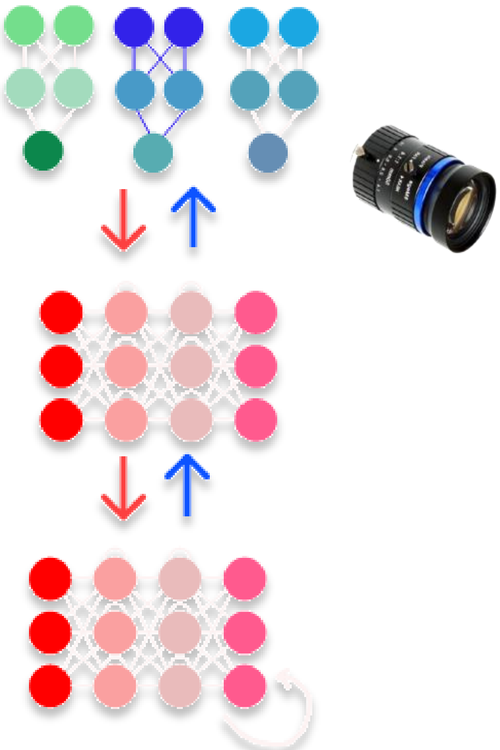
NX-414: Brain-like computation and intelligence

Martin Schrimpf

Lecture 5, March 19 2025

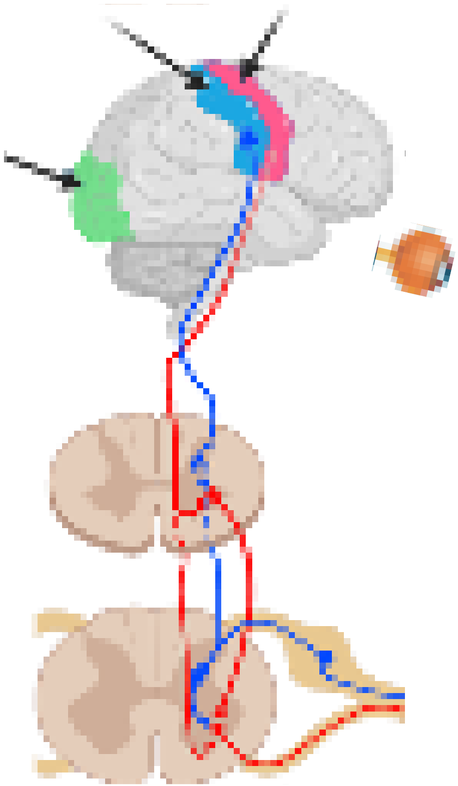
Biological Intelligence ⟷ Artificial Intelligence





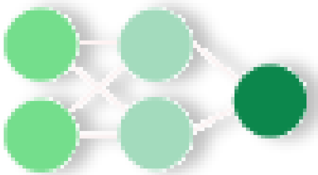
Hausmann & Marin-Vargas et al., 2021

Normative frameworks



Information theoretic

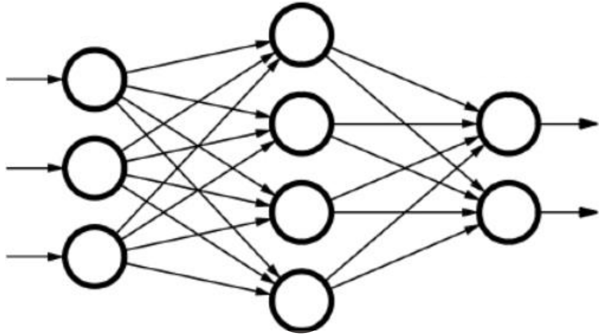
e.g. sparse coding,
redundancy reduction,
mutual information …

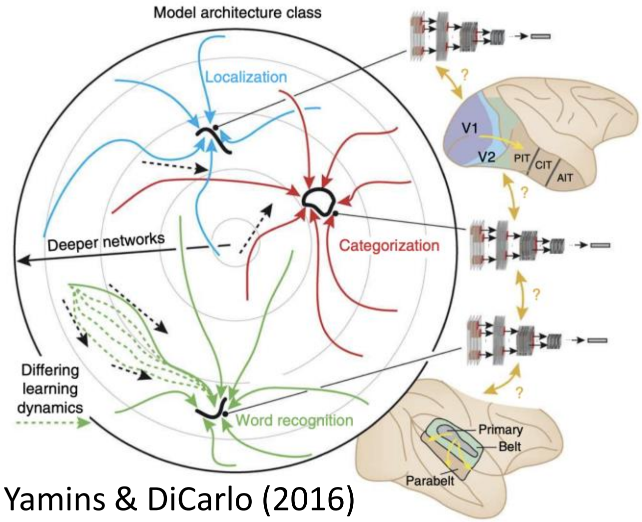


Utilitarian

e.g. **recognize objects**,
chase prey, navigate …

Using deep neural networks as goal-driven models of a system





Yamins & DiCarlo (2016)

<u>**Vision**</u>: object recognition.
Yamins & Hong et al. (2014), Schrimpf & Kubilius et al. (2018)

<u>Audition</u>: speech recognition, speaker & sound identification. Kell et al. (2018)

<u>Somatosentation</u>: shape recognition.
Zhuang et al. (2017)

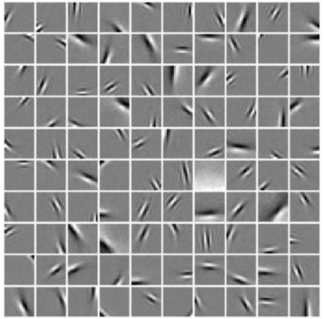
<u>Language</u>: next-word prediction.
Schrimpf et al. (2021)

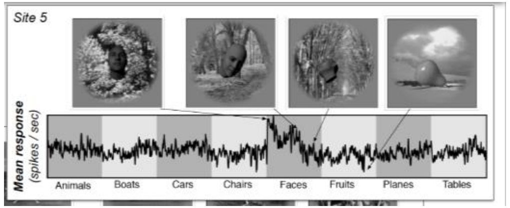
<u>Decision making</u>: context-dependent choice. Mante & Sussilo et al. (2013)

<u>Proprioception</u>: action recognition.
Sandbrink et al. (2023)

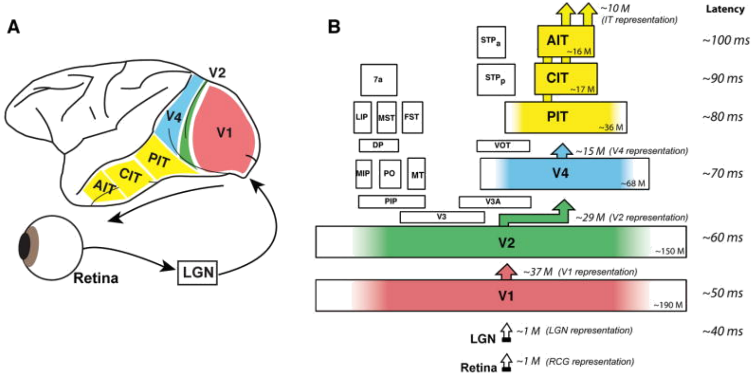
Recap from last week

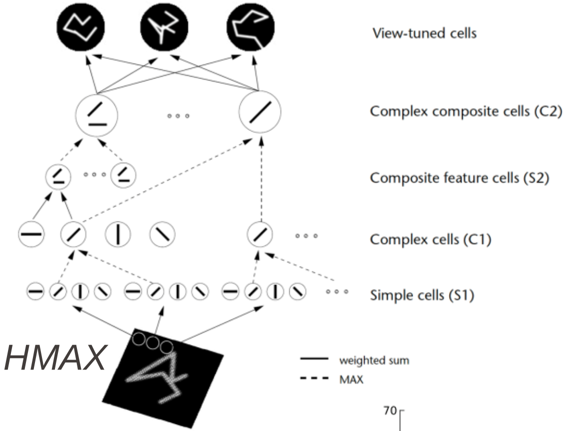
- Difficult for a long time to scale models beyond simple responses in V1



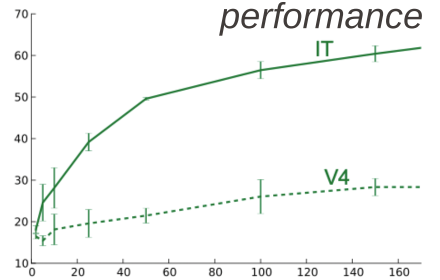


- IT –linear-readout→ object category



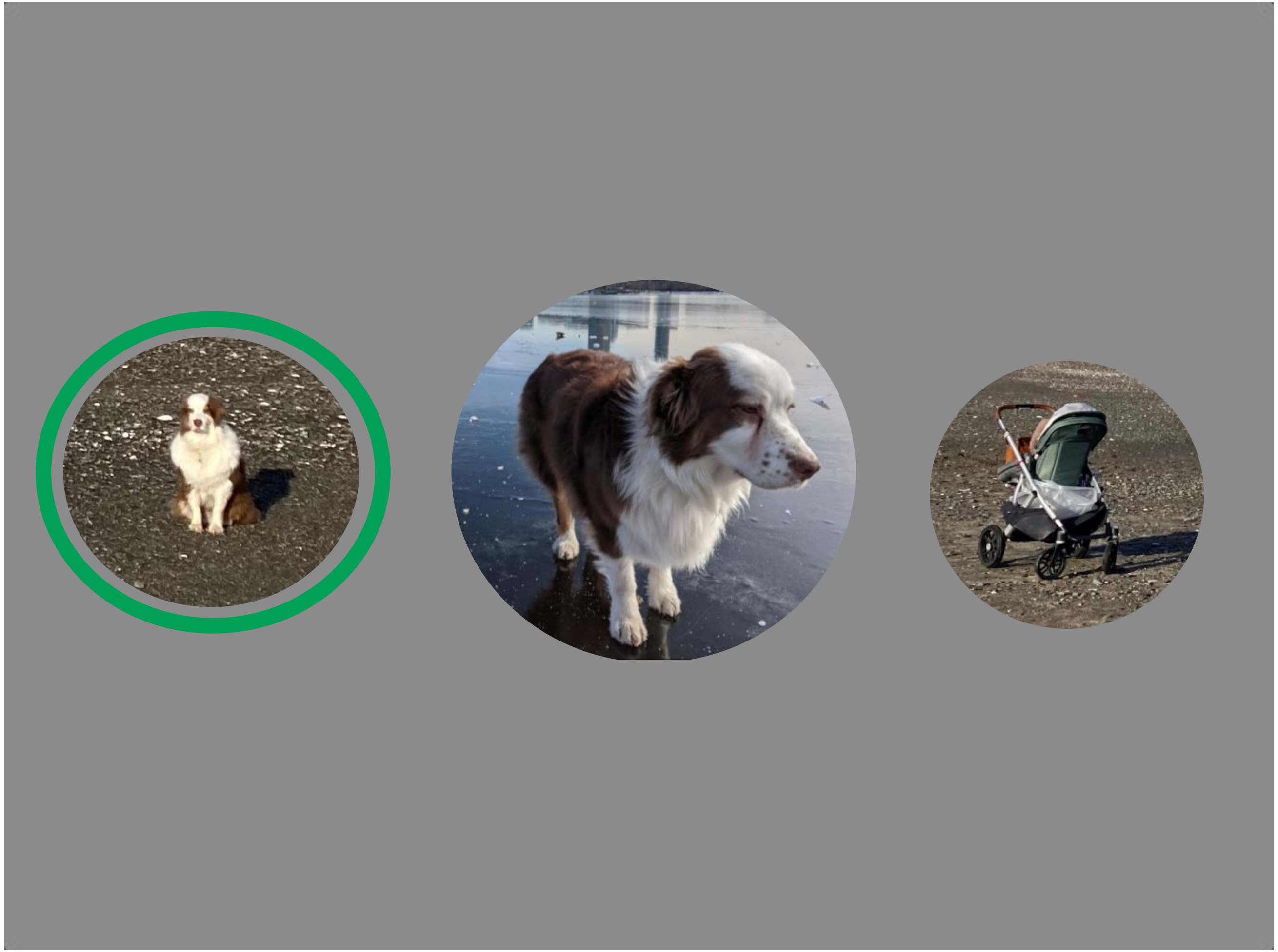


~human-level performance

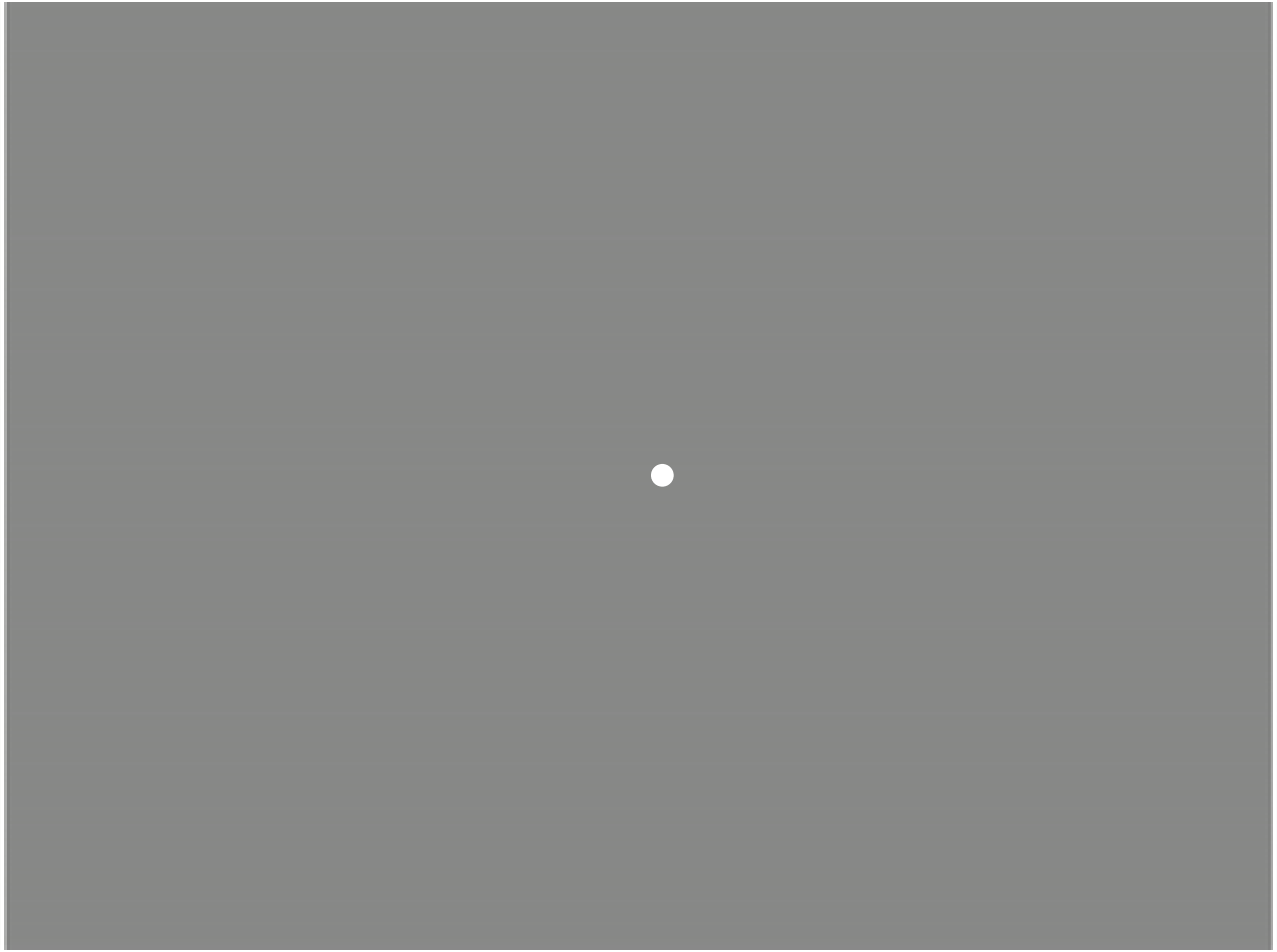


- Task optimization → brain-like internal representations

Behavioral experiment



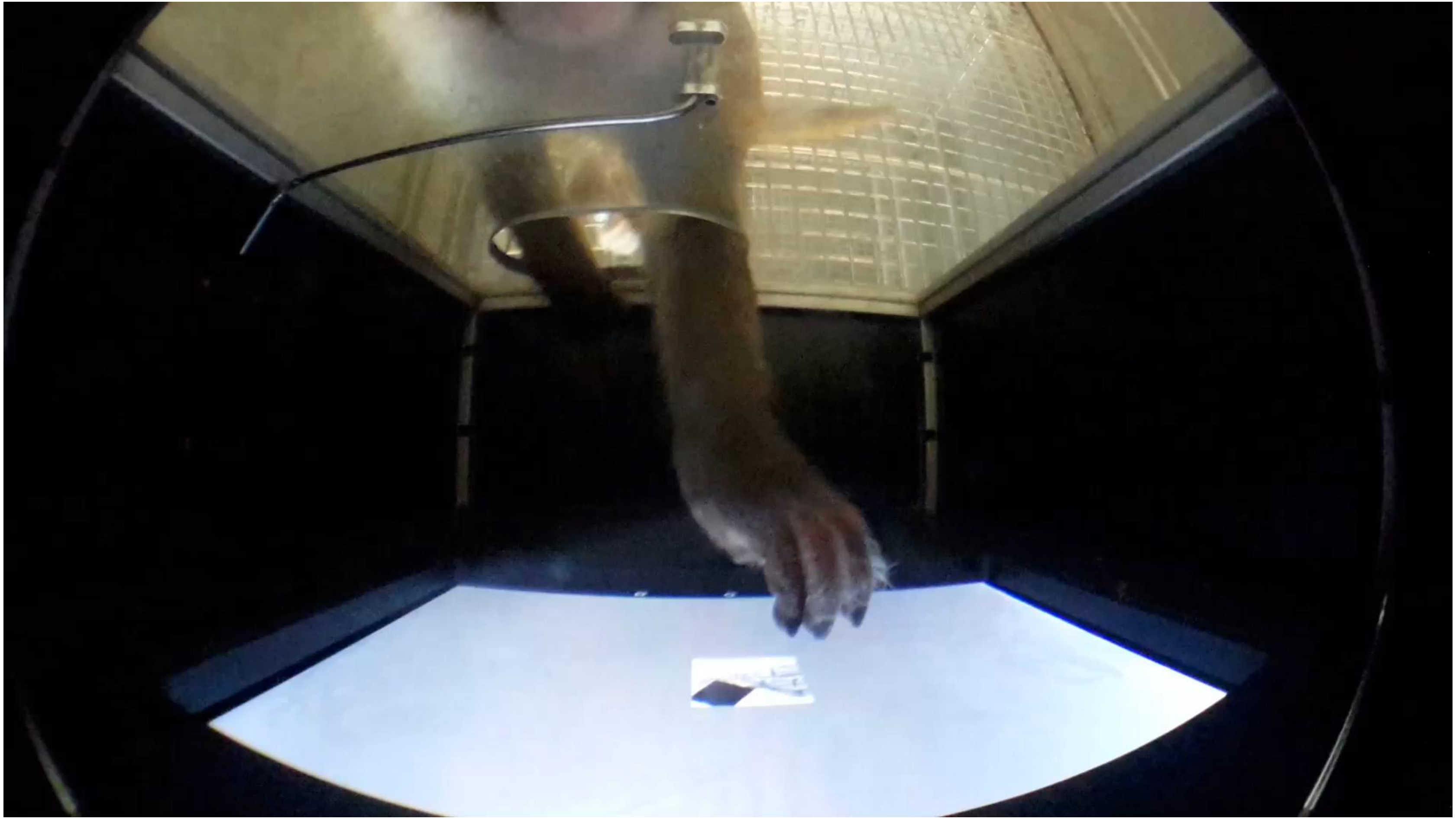
Behavioral experiment



Rajalingham, Issa*, et al. (JNeuro 2018)*

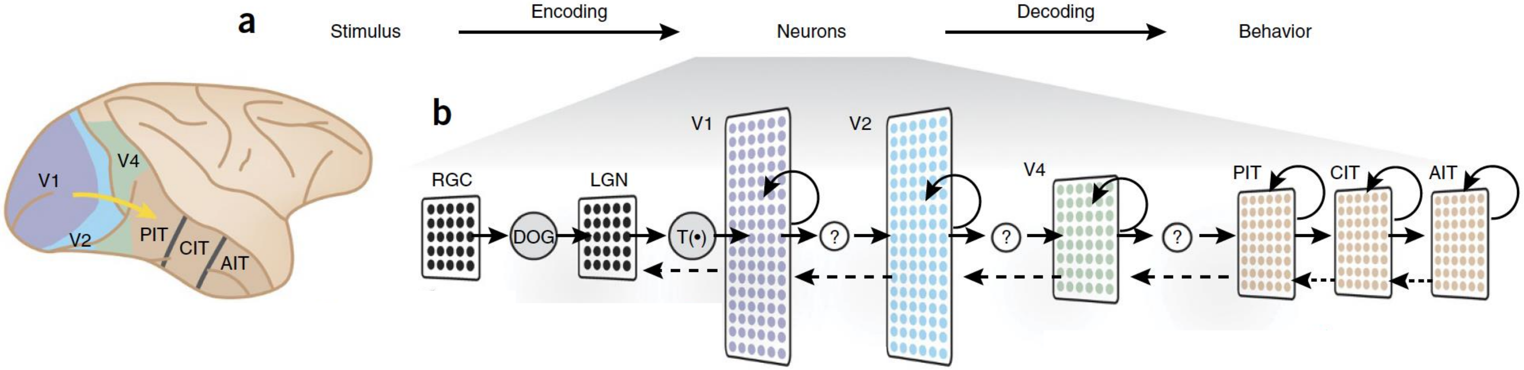
EPFL

Neural data



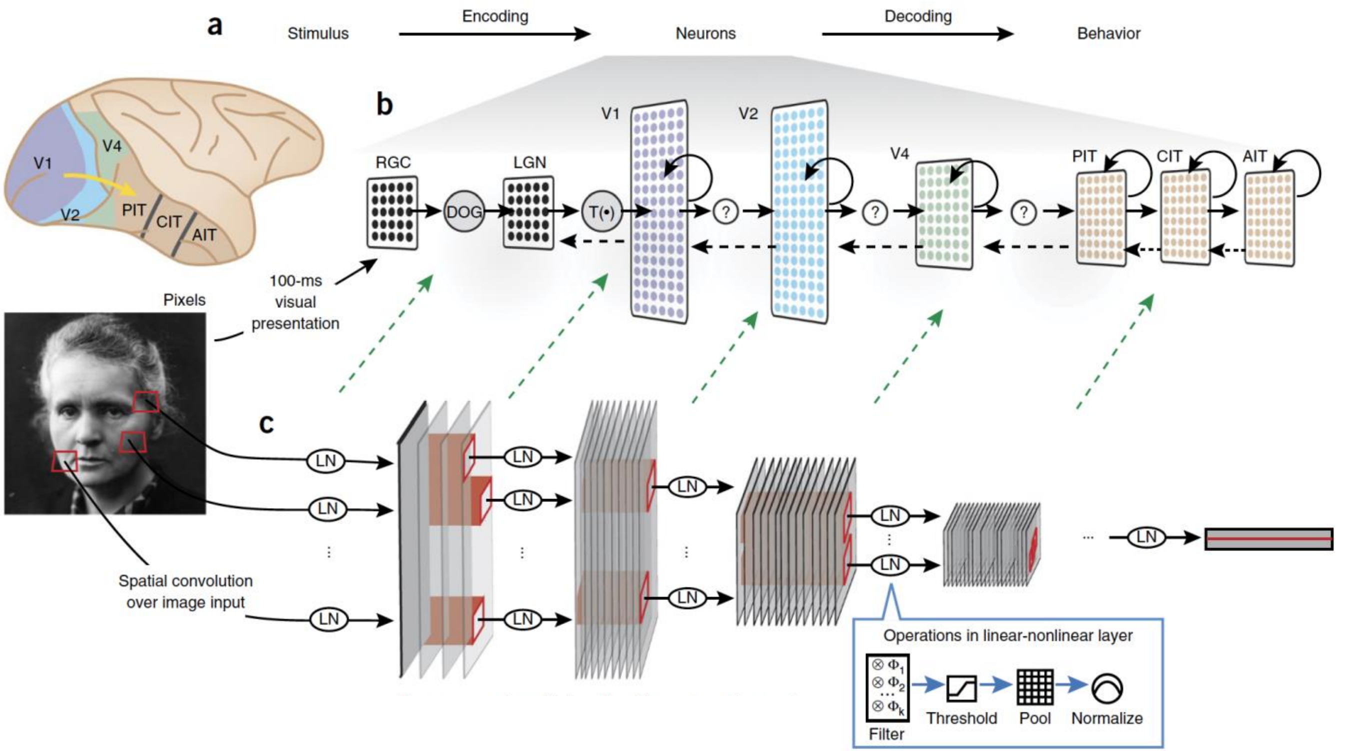
video courtesy of Kailyn Schmidt, MIT

Core-object recognition and the visual pathway



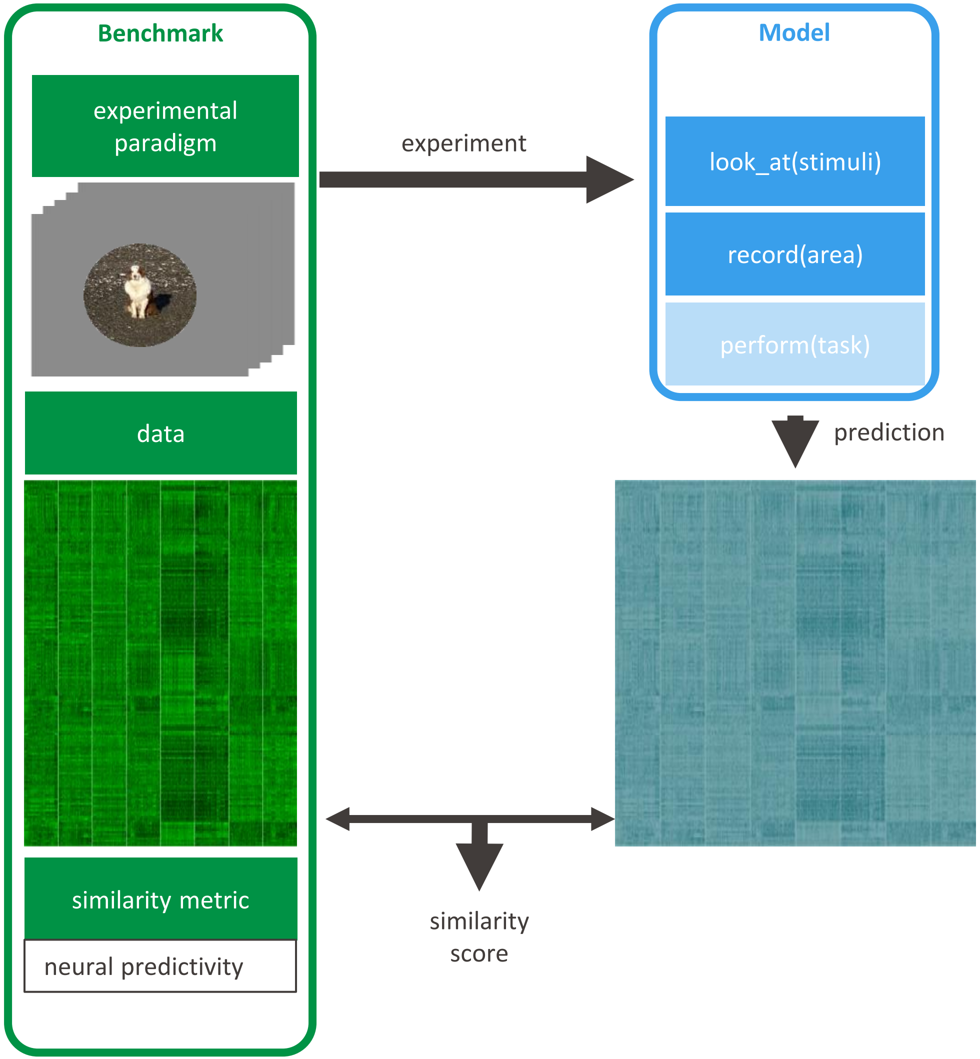
Yamins and DiCarlo, Nat Neuro 2016

Model building



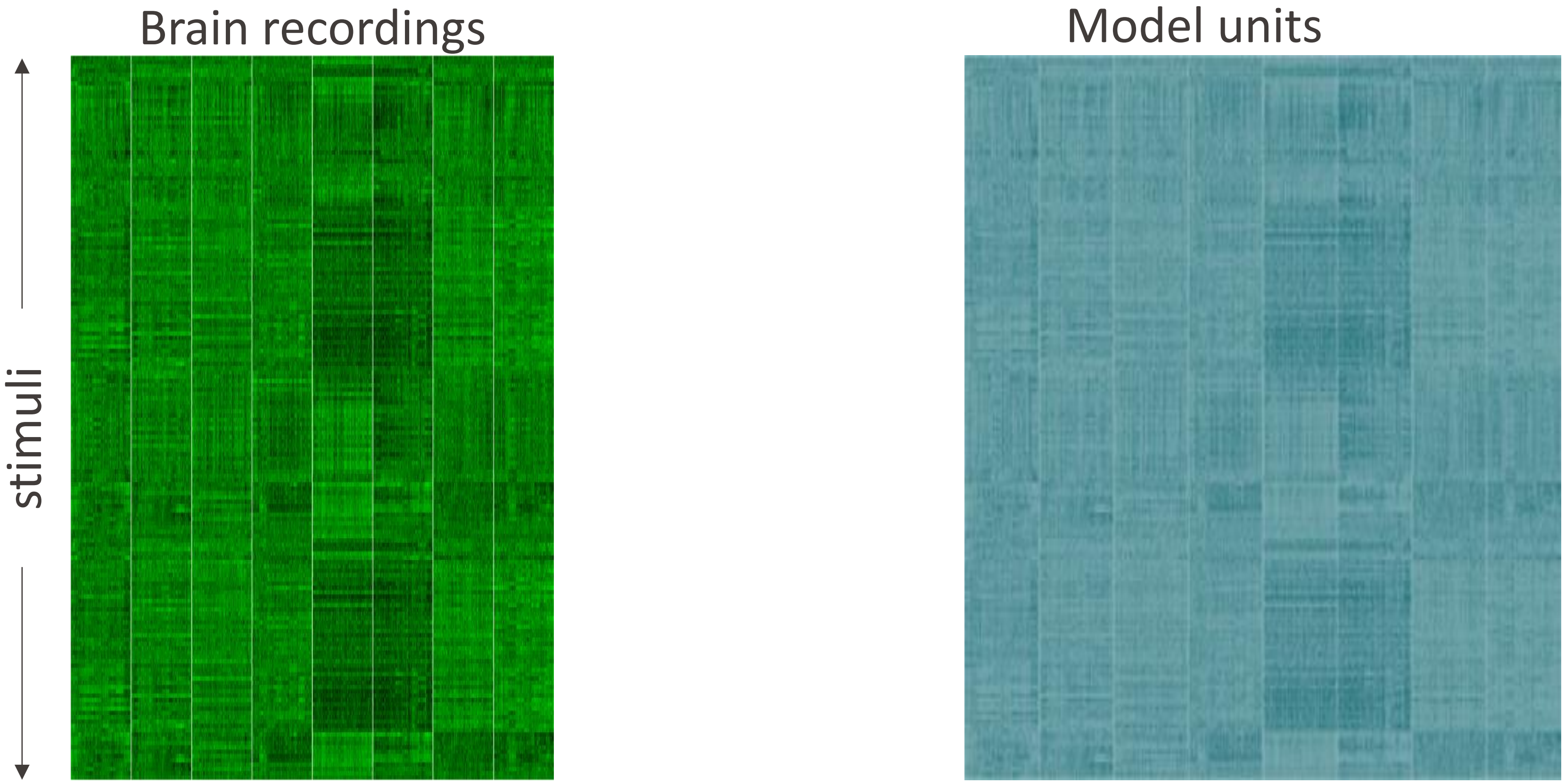
Yamins and DiCarlo, Nat Neuro 2016

Neural benchmark



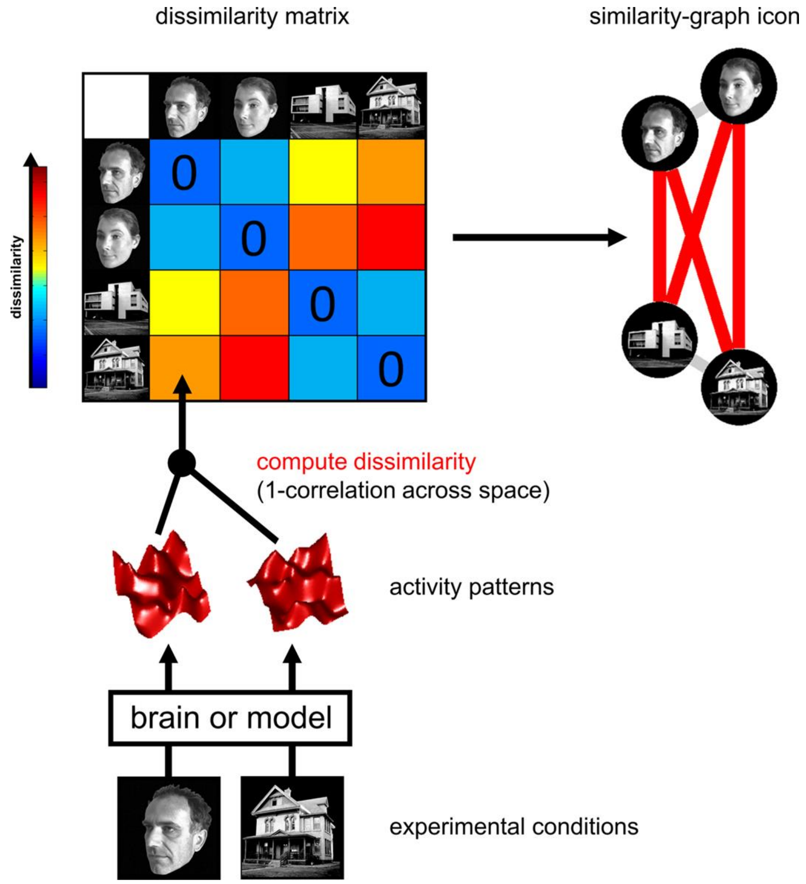
Neural benchmarks

Neural alignment = alignment between stimulus-matched recordings



Representational similarity analysis (RSA)

$$RDM(\phi(x))_{ij} = 1 - \frac{cov(\phi(x_i),\phi(x_j))}{\sqrt{var(\phi(x_i))\cdot var(\phi(x_j))}}$$



Kriegeskorte al. Frontiers in Systems Neuro 2008

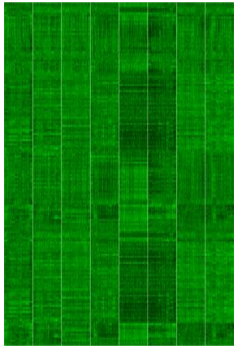
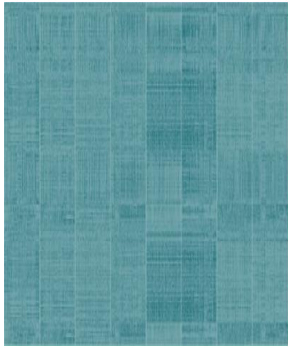
similarity metric

Representational similarity analysis (RSA)

Neural benchmarks

Brain recordings

Model units

score = corr(RDM () , RDM())

correlate upper triangular only

EPFL

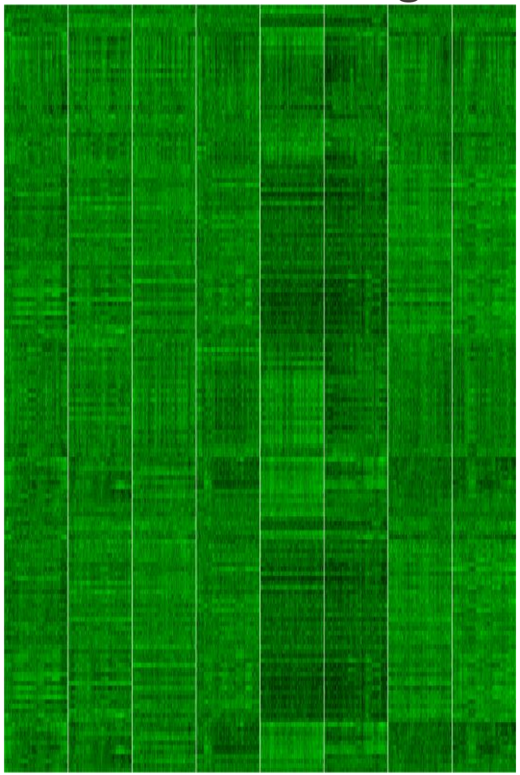
Neural benchmarks

similarity metric

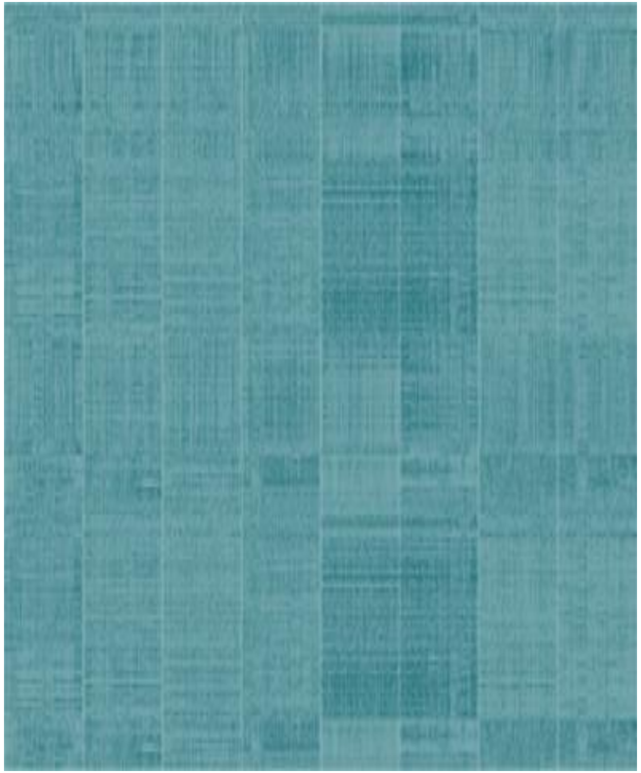
neural predictivity

Brain recordings

stimuli



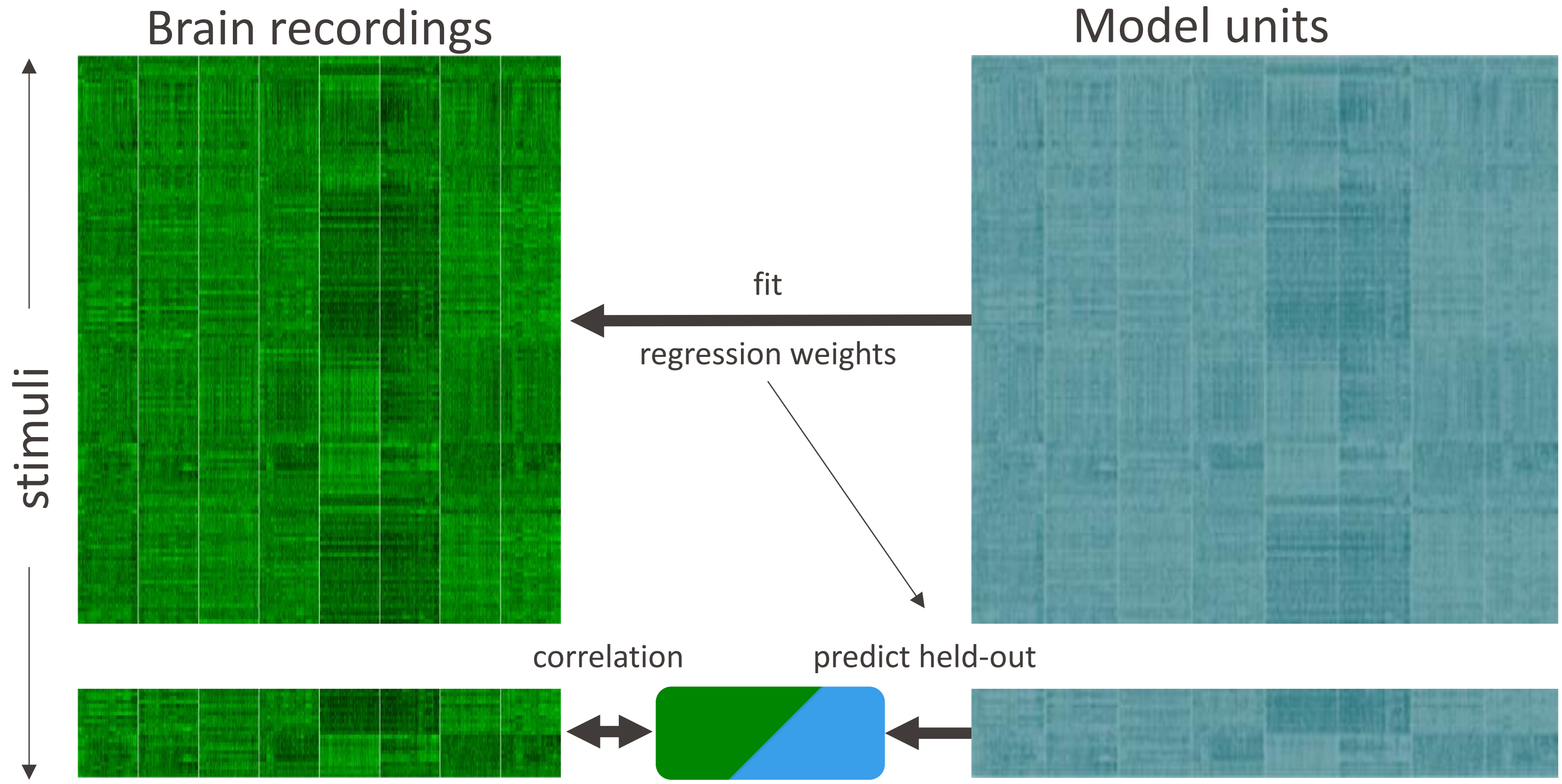
Model units



Neural benchmarks

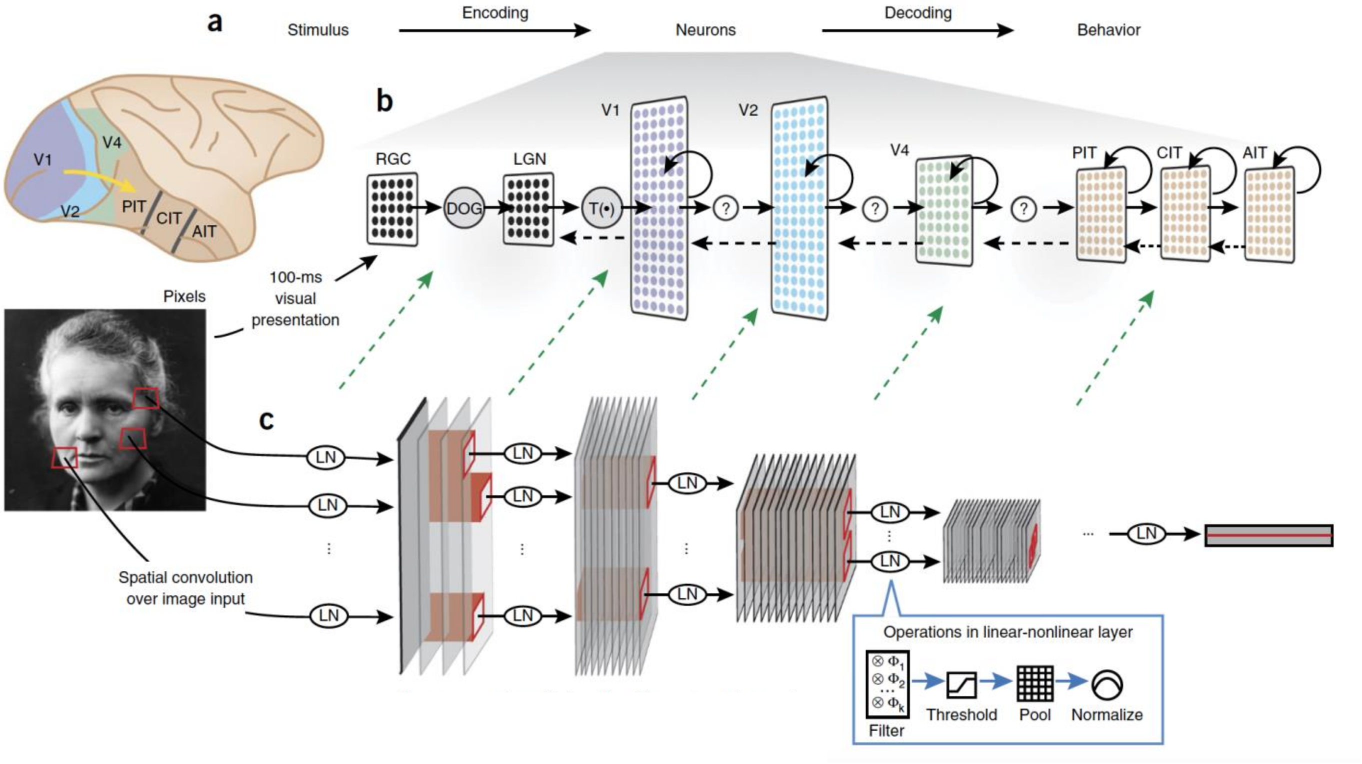
similarity metric

neural predictivity



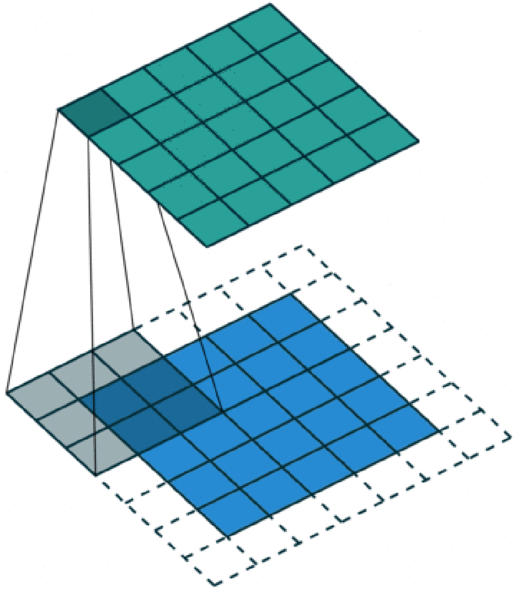
Yamins*, Hong*, et al. 2014 Schrimpf*, Kubilius*, et al. 2018

Model building

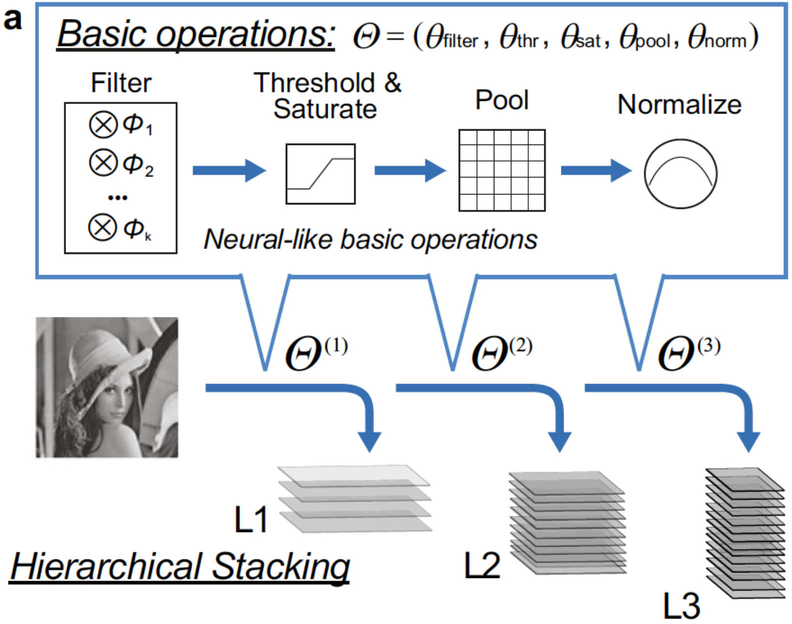


Yamins and DiCarlo, Nat Neuro 2016

Convolutions



+ pooling + hierarchy

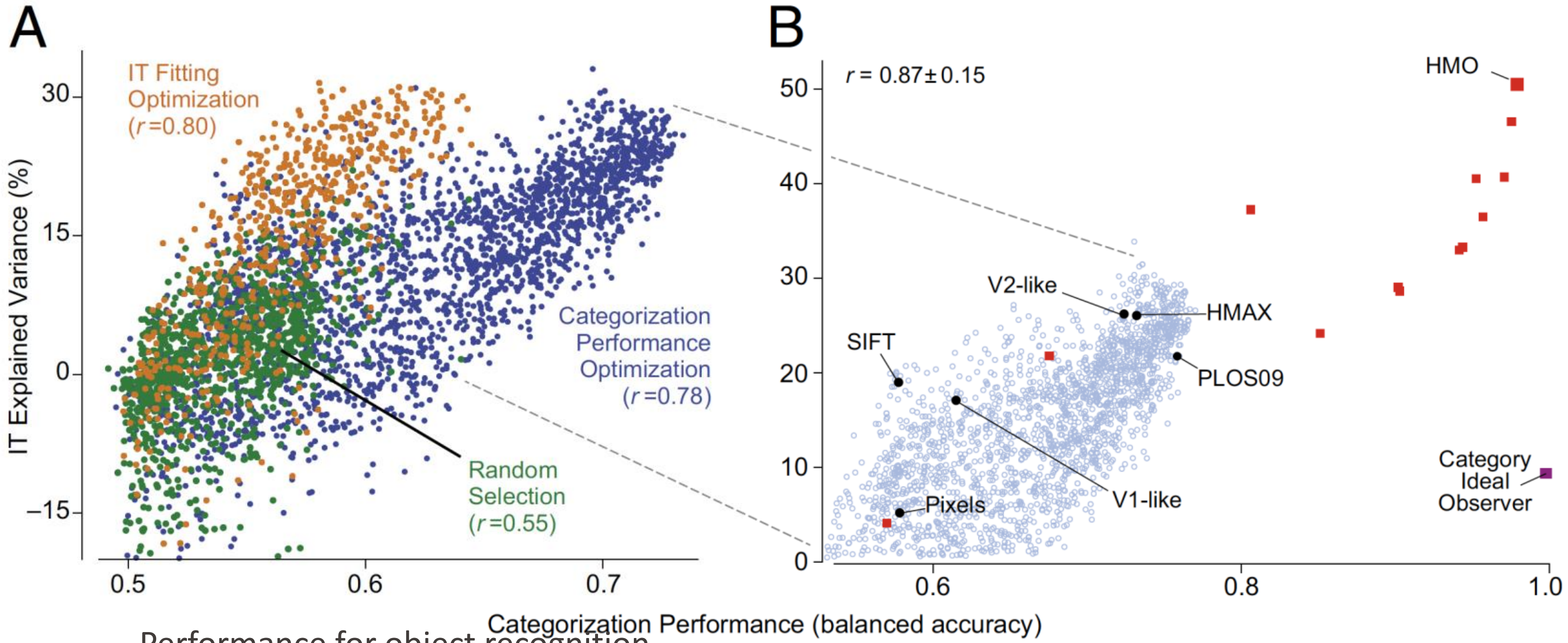


+ error-based reweighting (adaboost)
→ task optimization

via lots of compute
(Pinto et al. 2009; Bergstra et al. 2013)

https://github.com/vdumoulin/conv_arithmetic

Large scale architecture search and model comparison



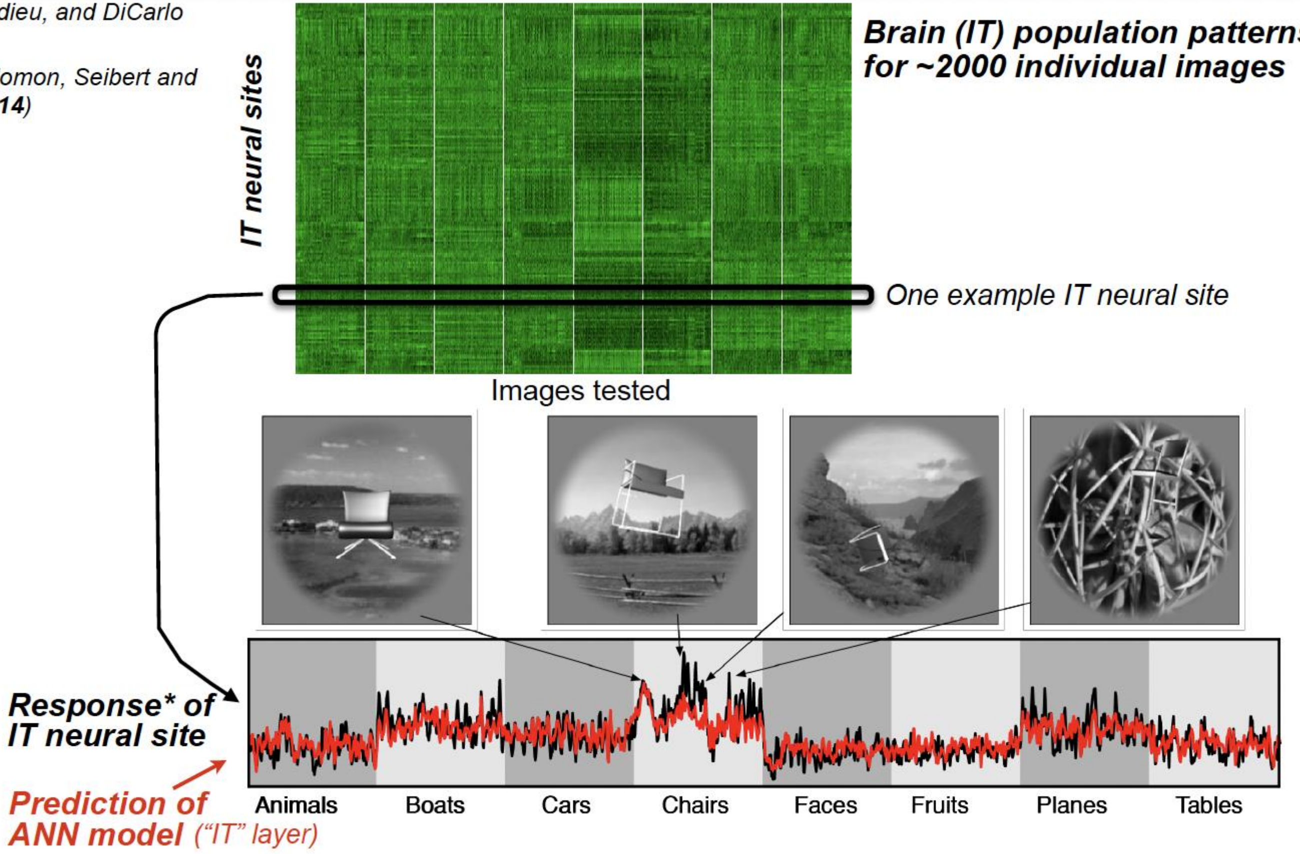
Performance for object recognition

Yamins & Hong et al., PNAS 2014

Task-trained models match brain's "hidden neurons"

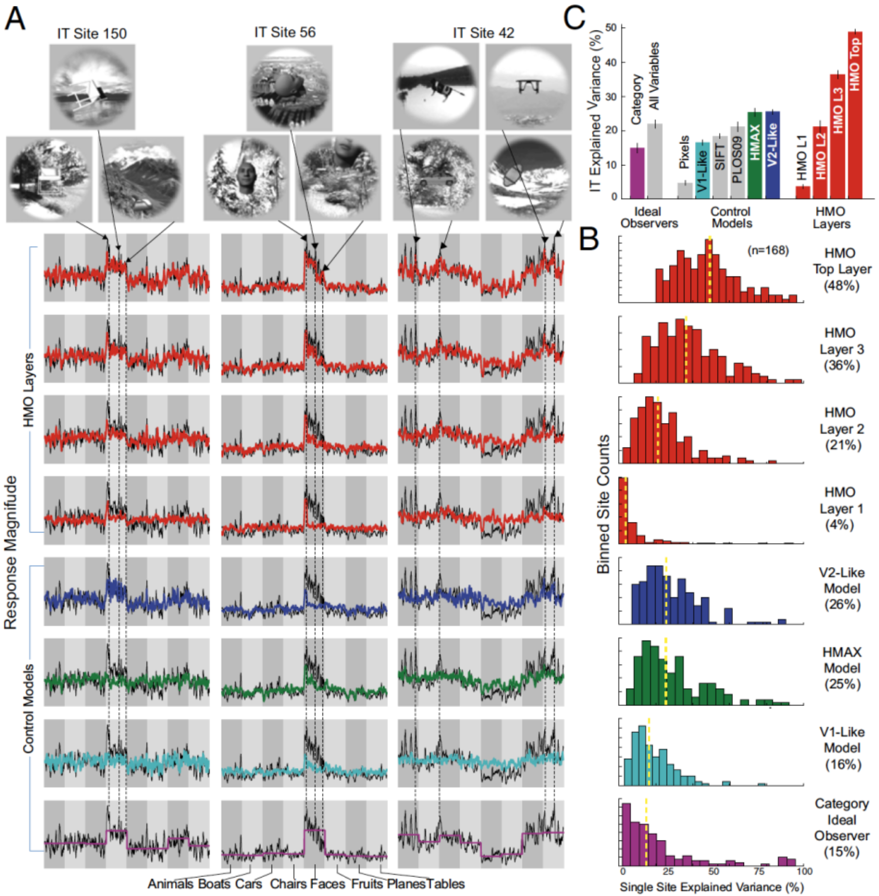
Yamins, Hong, Cadieu, and DiCarlo **NIPS (2013)**

Yamins, Hong, Solomon, Seibert and DiCarlo **PNAS (2014)**



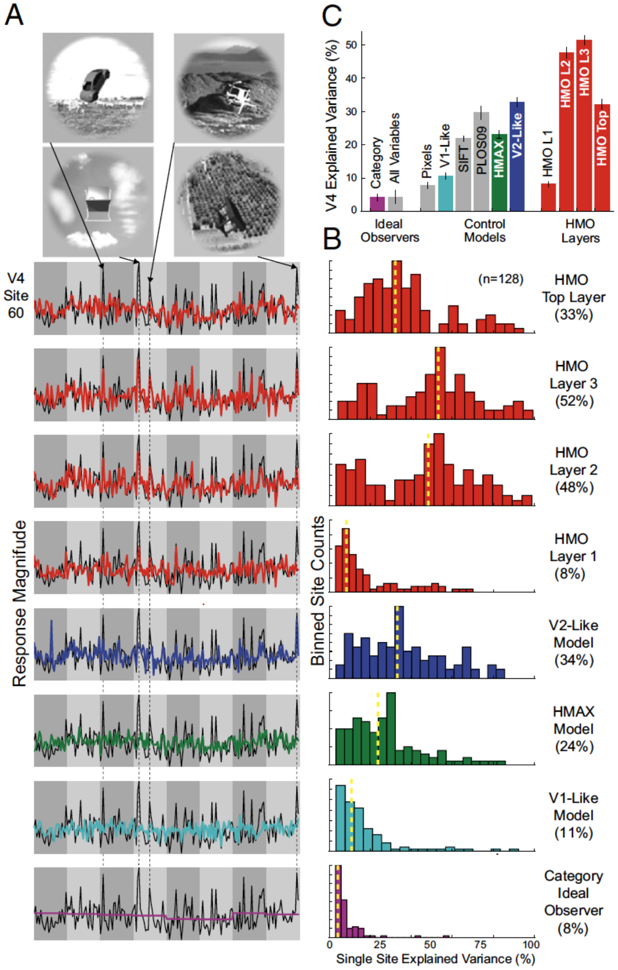
Slide from Jim DiCarlo, MIT

Predictions for IT



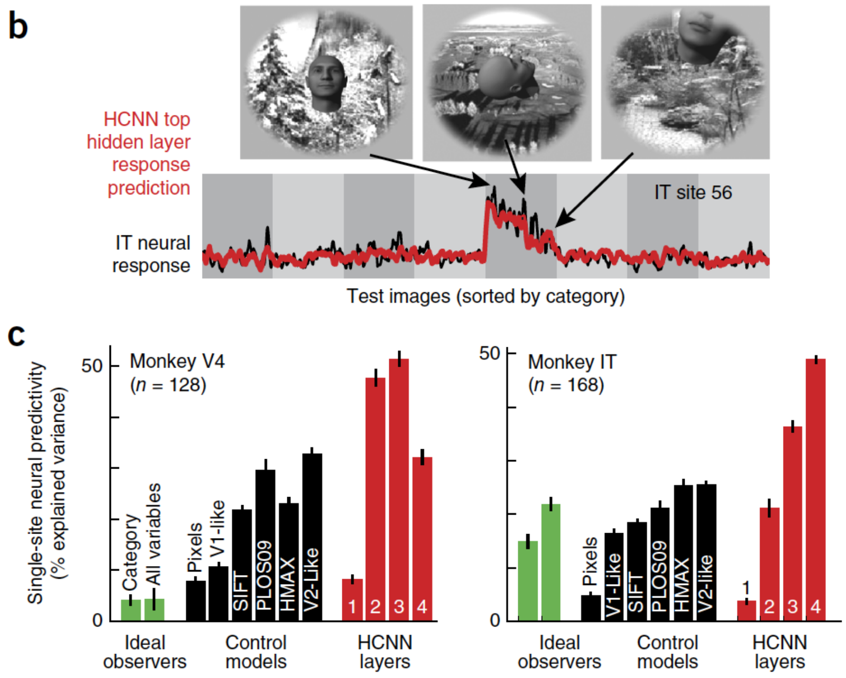
Yamins & Hong et al., PNAS 2014

Predictions for V4



Yamins & Hong et al., PNAS 2014

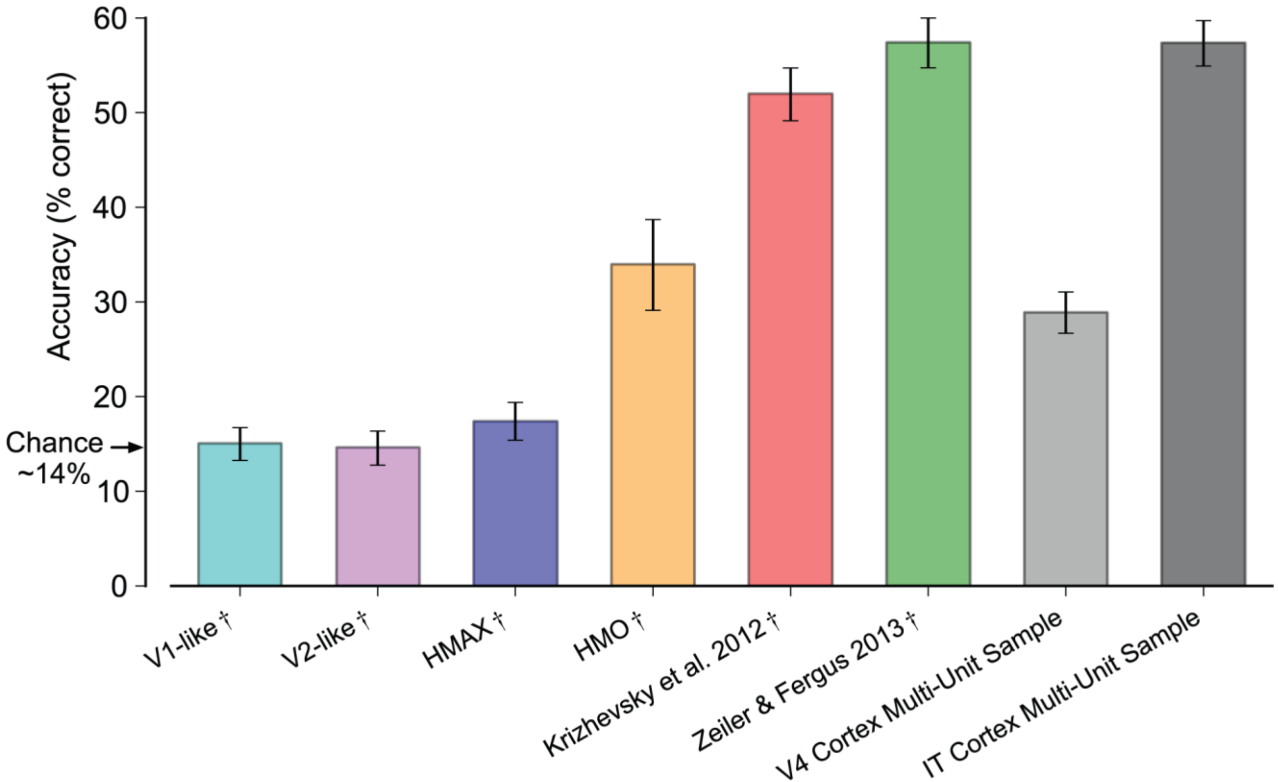
Hierarchy as a consequence of task-optimization



Yamins & Hong et al., PNAS 2014

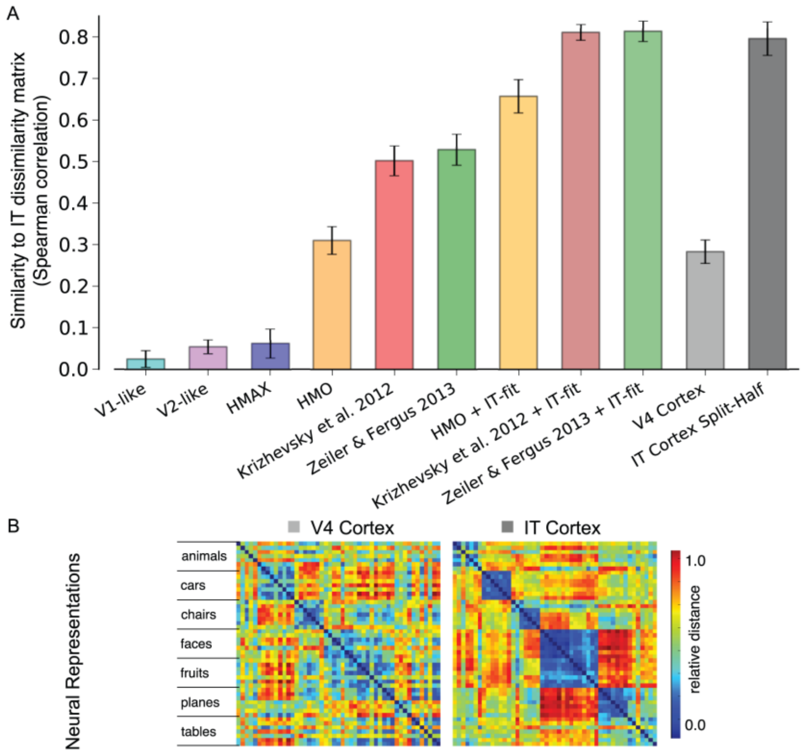
Category classification

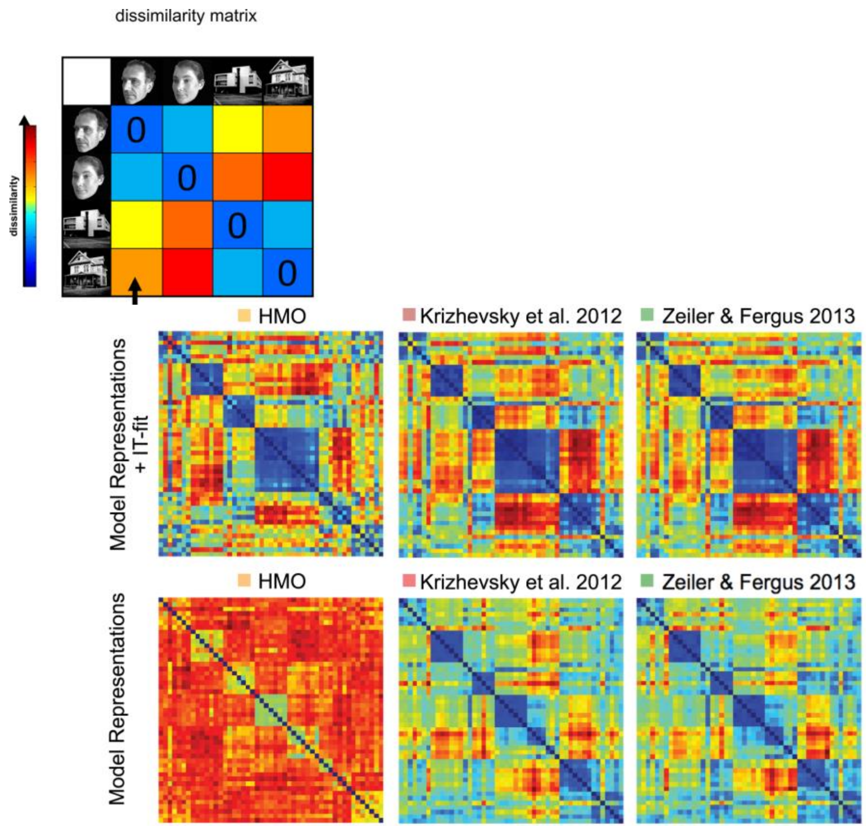
Comparing to backprop-trained deep nets…



Cadieu et al. PLOS CB 2014

Model comparison

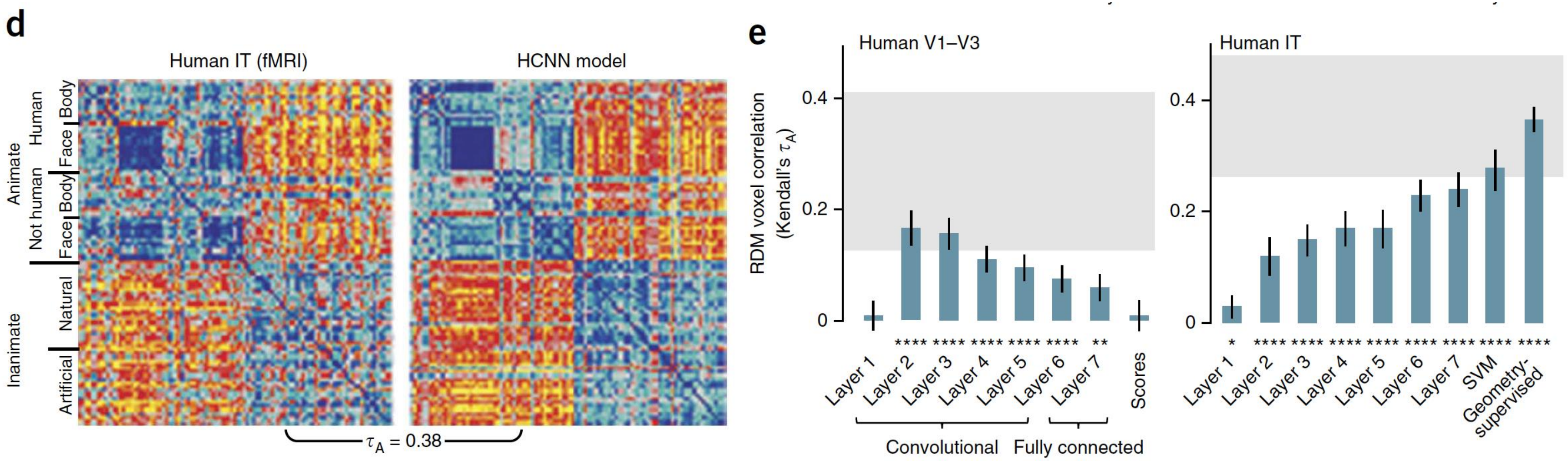




Backprop on ImageNet is much better than HMO.
Computer Vision – "as a side effect" – revolutionized modeling primate vision!

Cadieu et al. Plos CB 2014

fMRI based analysis in humans



Khaligh-Razavi and Kiegeskorte PloS CB 2014

- Work on 12 exciting projects
- Develop new skills
- Collaborate with like-minded people
- Get expert mentorship
- Get inspired by talks from Alumni
- Pitch projects and win prizes
- Have fun during your Easter break

Lemanic Life Sciences Hackathon

April 23 - 25, 2025

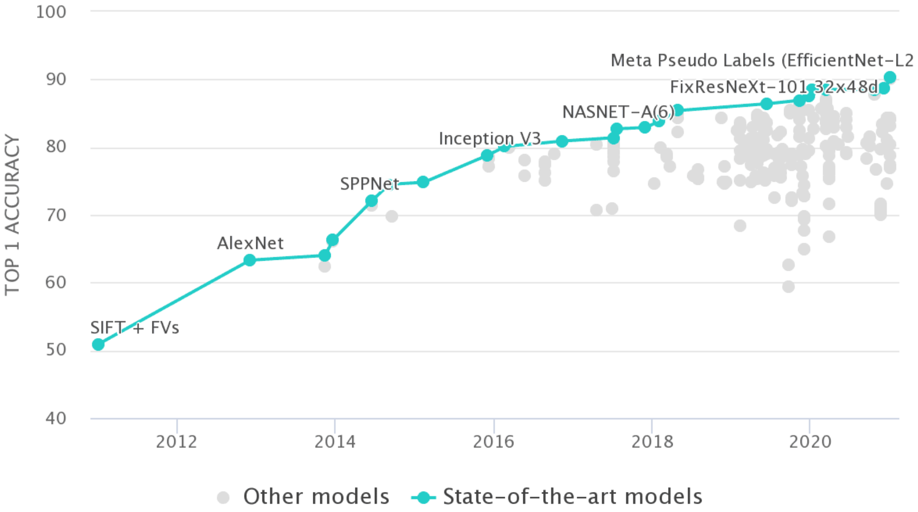
Hall SV and SV 1717

ImageNet challenge

1000 classes

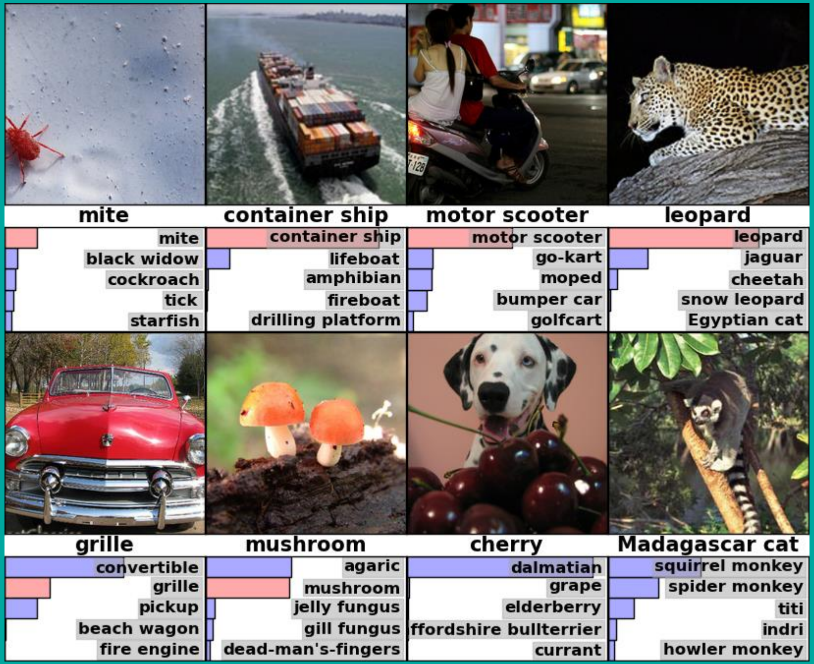
1.2 training images

150,000 test images



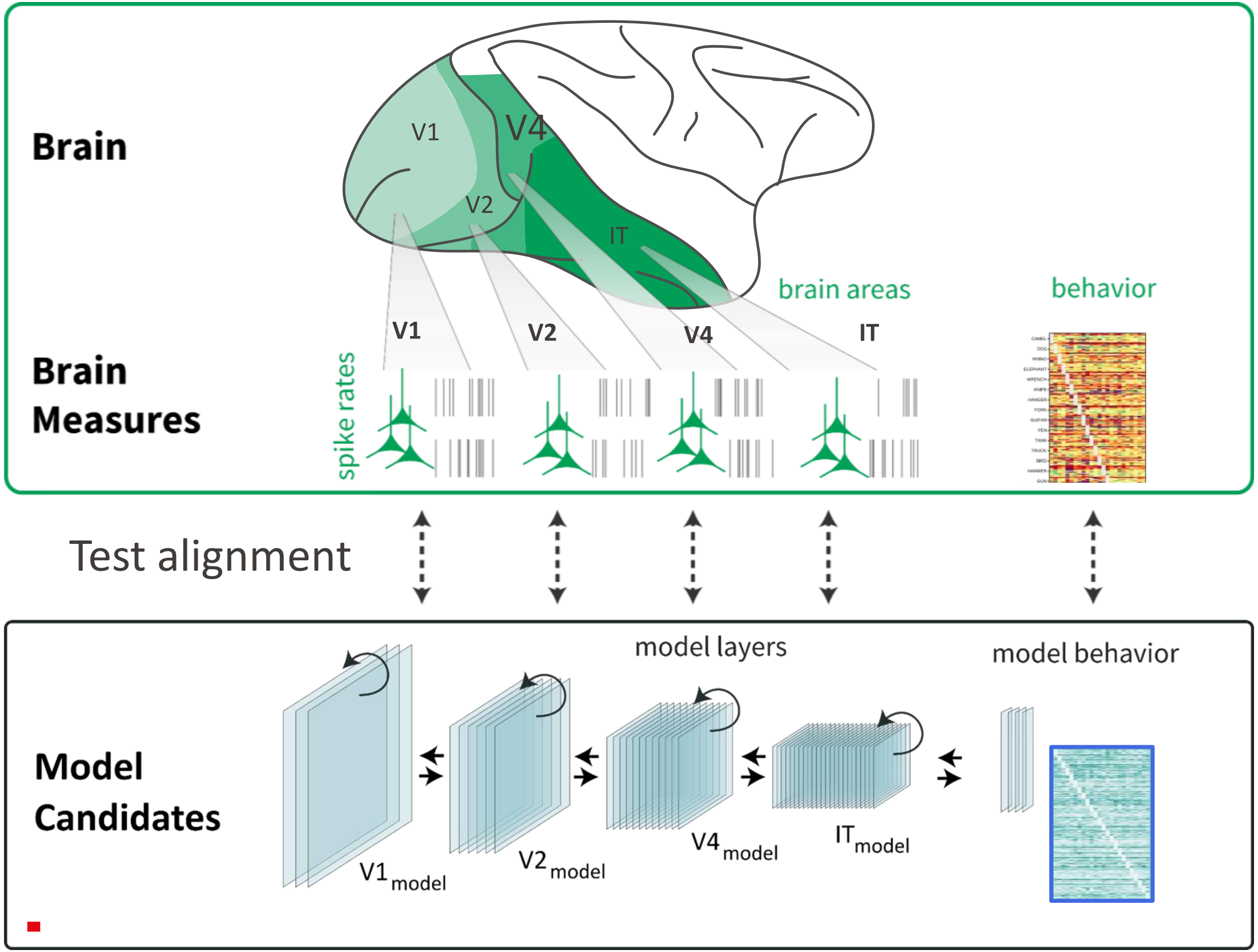
From papers with code

AlexNet predictions



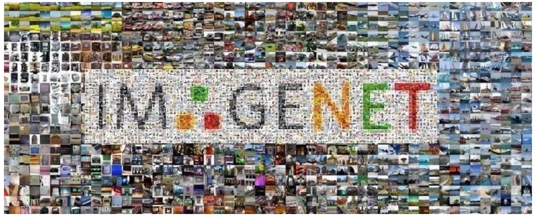
Krizhevsky, Sutskever, Hinton, NeurIPS 2012

Integrative testing of models on brain + behavioral data



Artificial neural network models

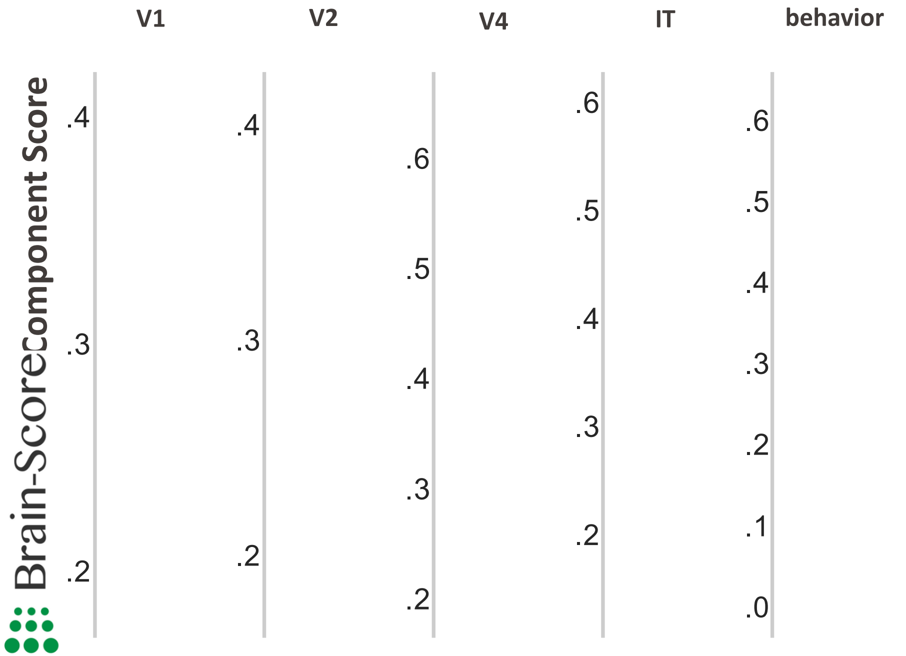
- Trained for computational task, weights optimized via backprop



- Internal processing stages (hidden layers, “deep” learning)
- Accept any new input (pixels)

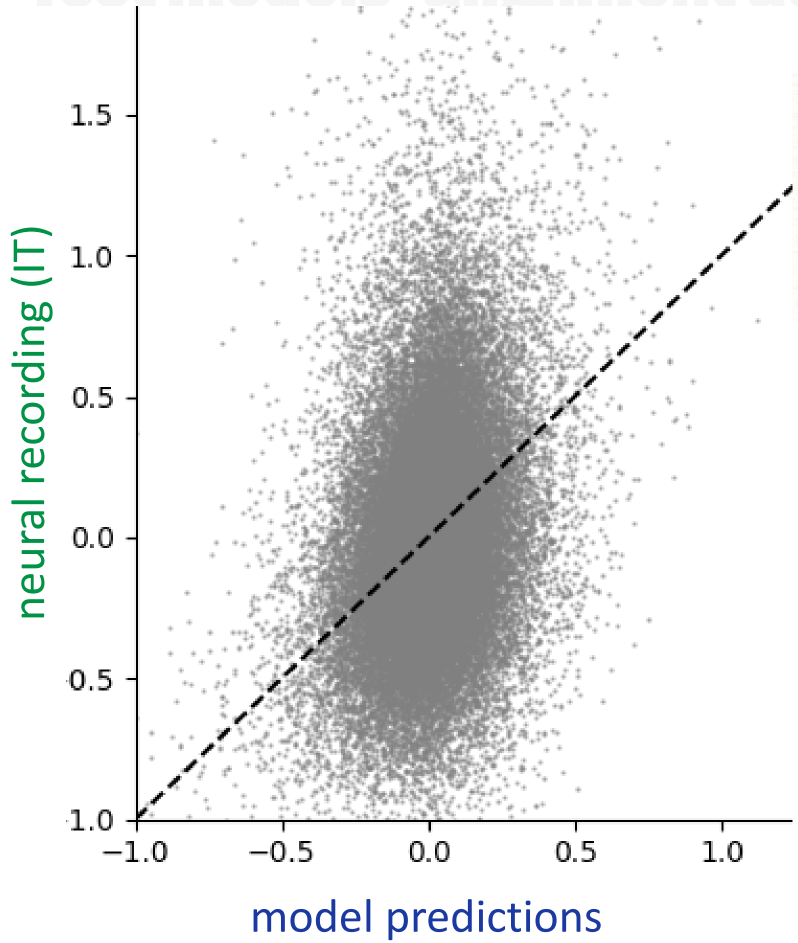
EPFL

Test models' alignment across visual ventral stream

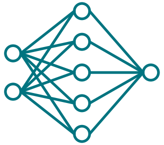


Schrimpf, Kubilius*, et al. (2018)*
V1, V2 data: Freeman, Ziemba*, et al. (2013)*
V4, IT data: Majaj, Hong*, et al. (2015)*
behavioral data: Rajalingham, Issa*, et al. (2018)*

Test models' alignment across visual ventral stream



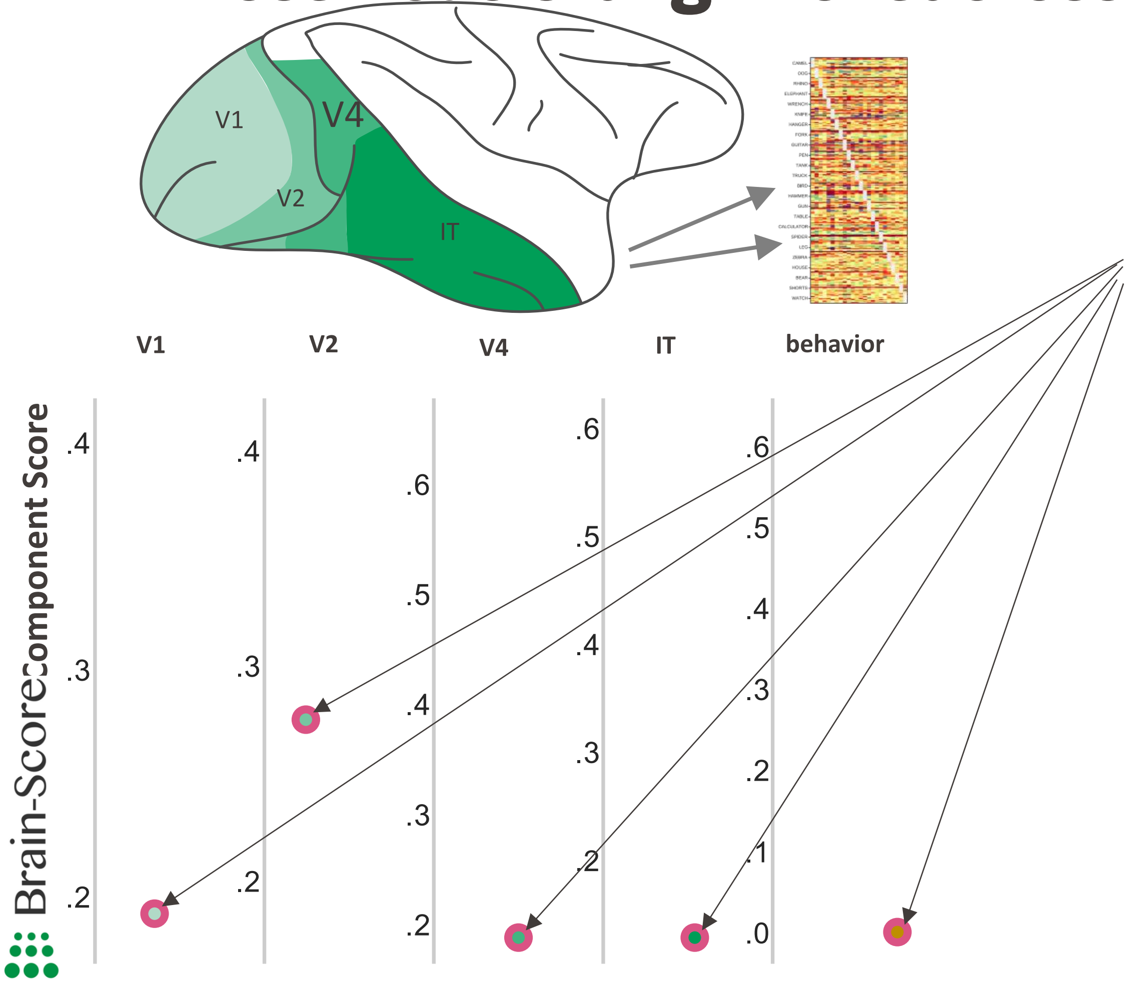
Model candidates tested:



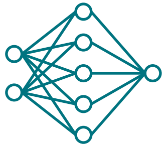
hmax *classic neuroscience model*

Schrimpf, Kubilius*, et al. (2018)*
V1, V2 data: Freeman, Ziemba*, et al. (2013)*
V4, IT data: Majaj, Hong*, et al. (2015)*
behavioral data: Rajalingham, Issa*, et al. (2018)*

Test models' alignment across visual ventral stream



Model candidates tested:

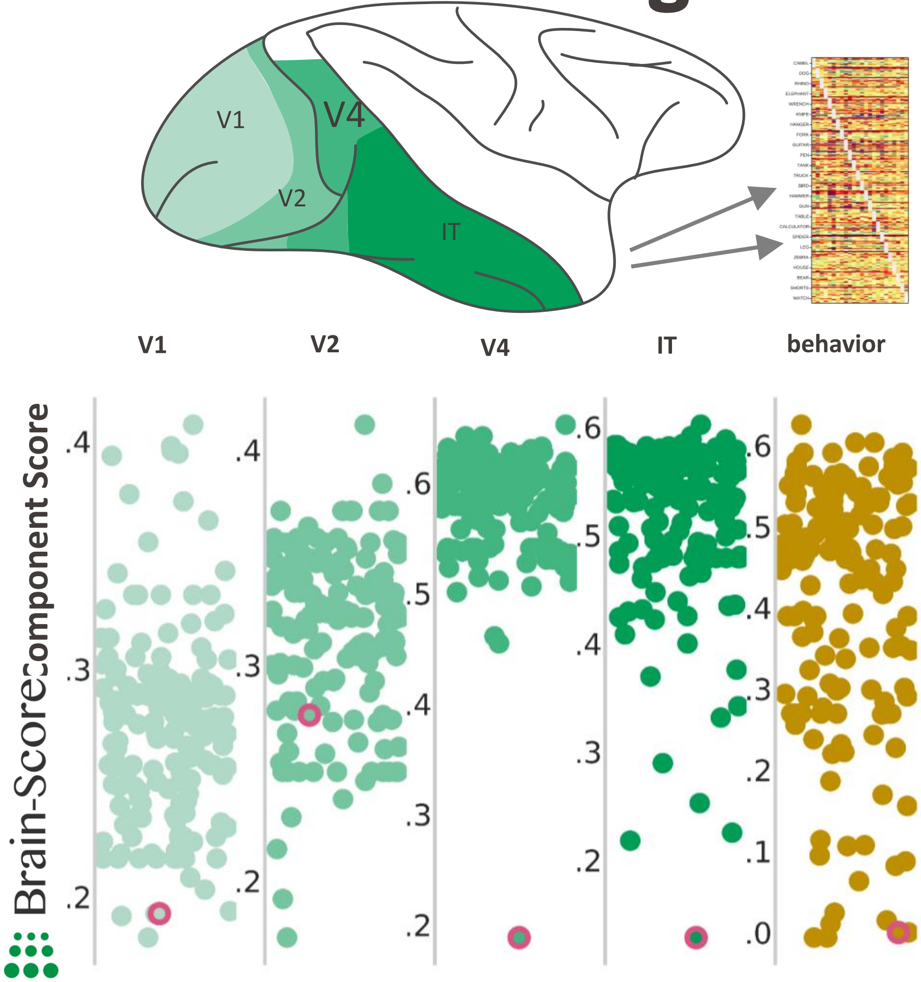


hmax *classic neuroscience model*

Schrimpf*, Kubilius*, et al. (2018)
V1, V2 data: Freeman, Ziemba*, et al. (2013)*
V4, IT data: Majaj, Hong*, et al. (2015)*
behavioral data: Rajalingham, Issa*, et al. (2018)*

EPFL

Test models' alignment across visual ventral stream



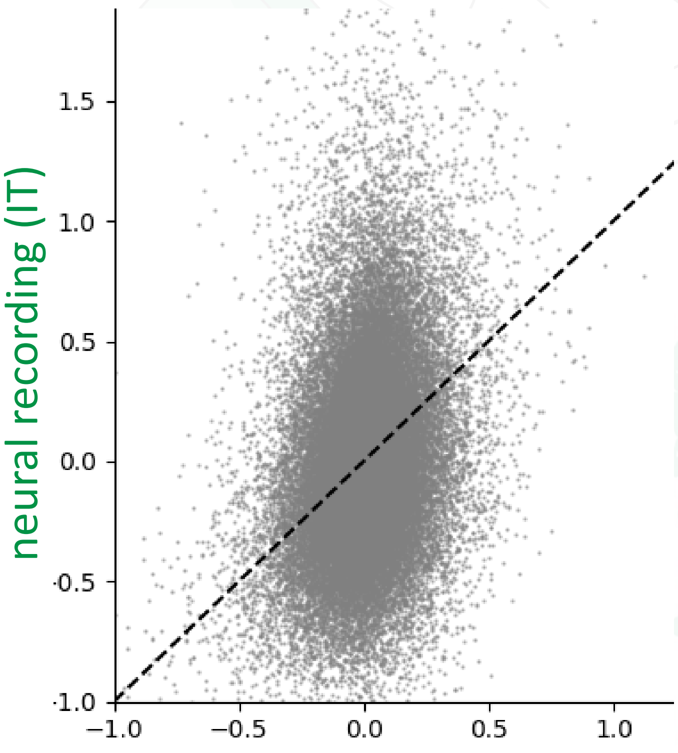
Model candidates tested:

ML models

hmax	nasnet_large
vgg-16	nasnet_mobile
vgg-19	pnasnet_large
densenet-121	resnet-101_v1
densenet-169	resnet-101_v2
densenet-201	resnet-152_v1
inception_resnet_v2	resnet-152_v2
inception_v1	resnet-18
inception_v2	resnet-34
inception_v3	resnet-50_v1
inception_v4	resnet-50_v2
mobilenet_v1_0.25_128	squeezenet1_0
mobilenet_v1_0.25_160	squeezenet1_1
mobilenet_...	xception
mobilenet_v2_1.3_224	...
mobilenet_v2_1.4_224	

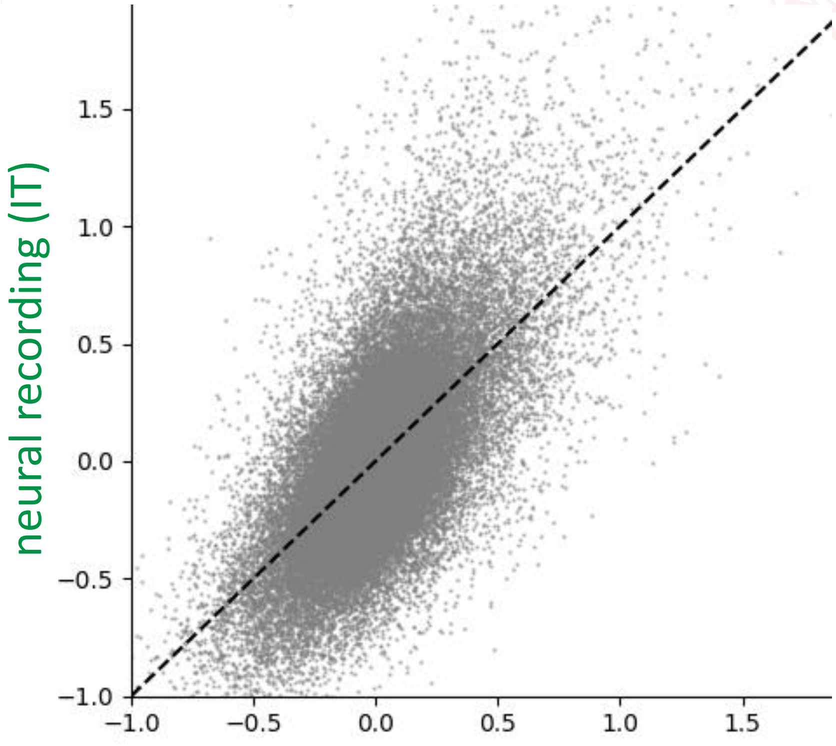
Schrimpf, Kubilius*, et al. (2018)*
V1, V2 data: Freeman, Ziemba*, et al. (2013)*
V4, IT data: Majaj, Hong*, et al. (2015)*
behavioral data: Rajalingham, Issa*, et al. (2018)*

hmax



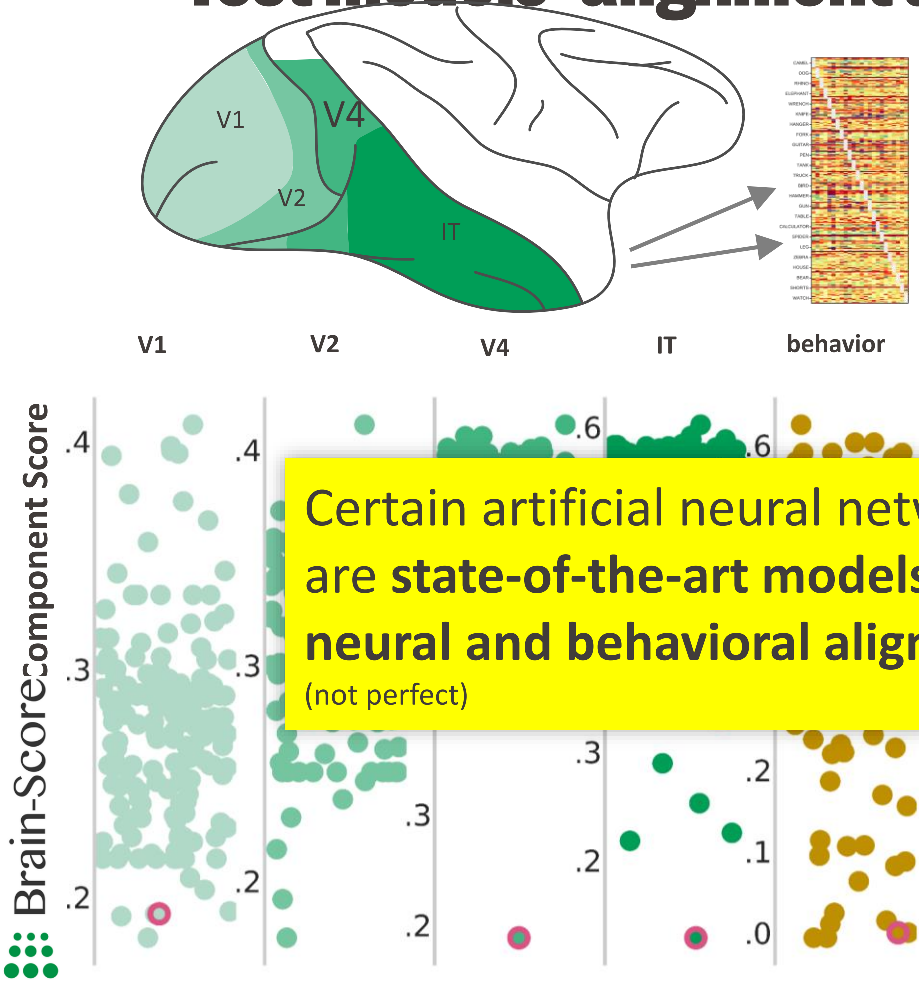
model predictions

resnext-wsl



model predictions

Test models' alignment across visual ventral stream



Certain artificial neural networks are **state-of-the-art models of neural and behavioral alignment**
(not perfect)

Model candidates tested:

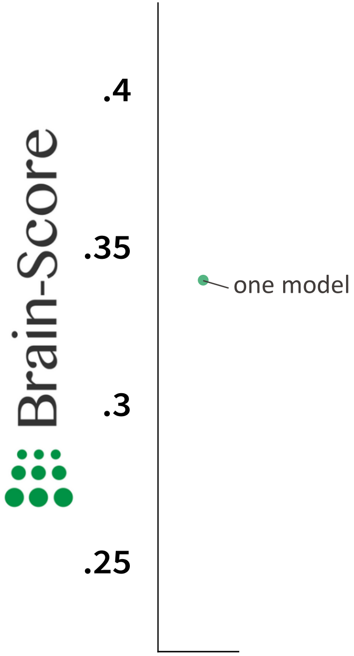
hmax
vgg-16
vgg-19
densenet-121
densenet-
densenet-
inception_
inception_
inception_
inception_
inception_
mobilenet
mobilenet_
mobilenet_...
mobilenet_v2_1.3_224
mobilenet_v2_1.4_224

nasnet_large
nasnet_mobile
pnasnet_large
resnet-101_v1
xception
...

All Computer Vision models are **trained on a task** without biological data

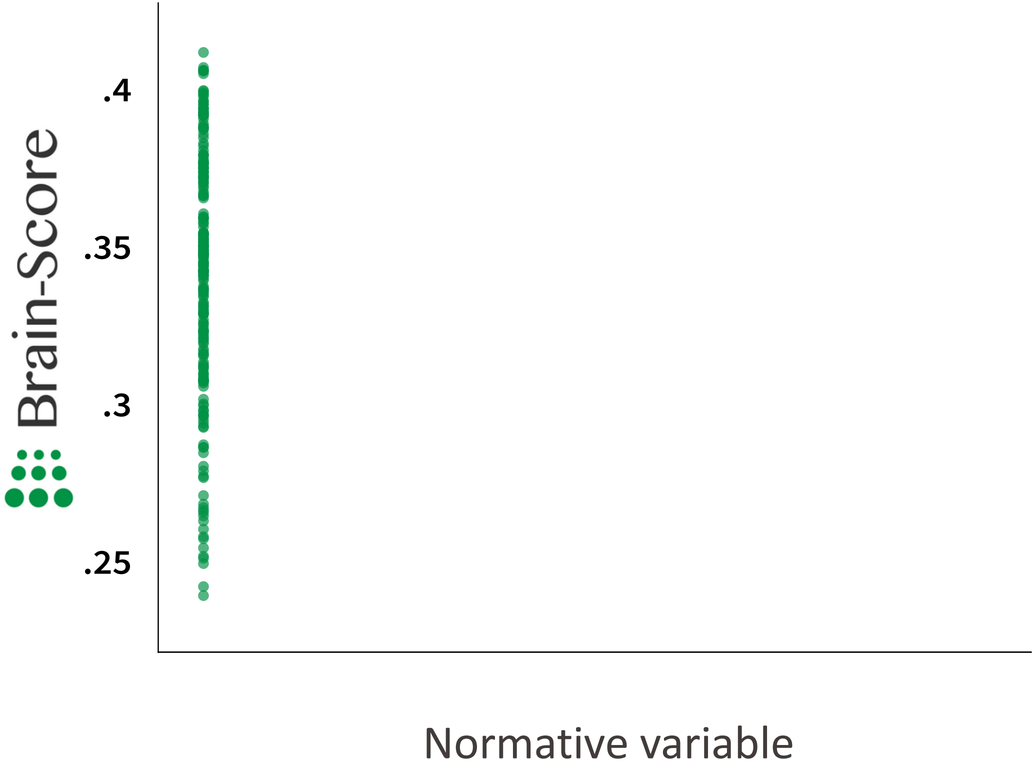
Schrimpf, Kubilius*, et al. (2018)*
V1, V2 data: Freeman, Ziemba*, et al. (2013)*
V4, IT data: Majaj, Hong*, et al. (2015)*
behavioral data: Rajalingham, Issa*, et al. (2018)*

What explains the model differences?



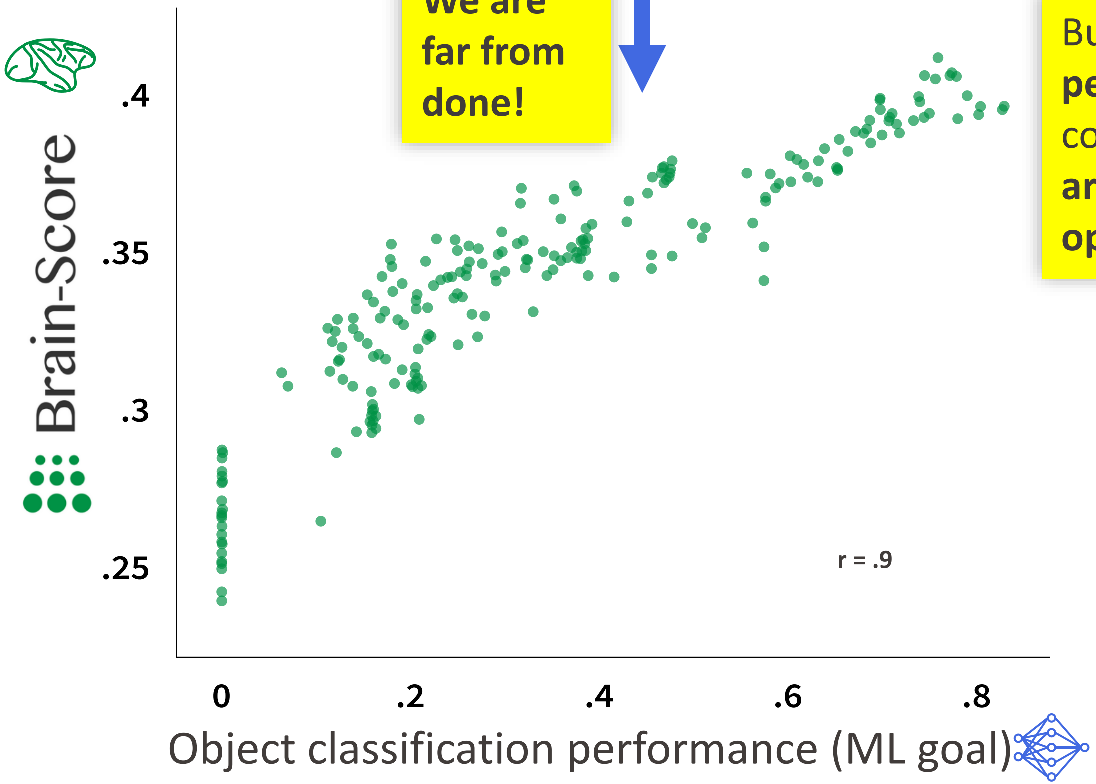
Schrimpf*, Kubilius*, et al. 2018

What explains the model differences?



cf. Yamins, Hong*, et al. 2014*
Schrimpf, Kubilius*, et al. 2018*

What explains the model differences?



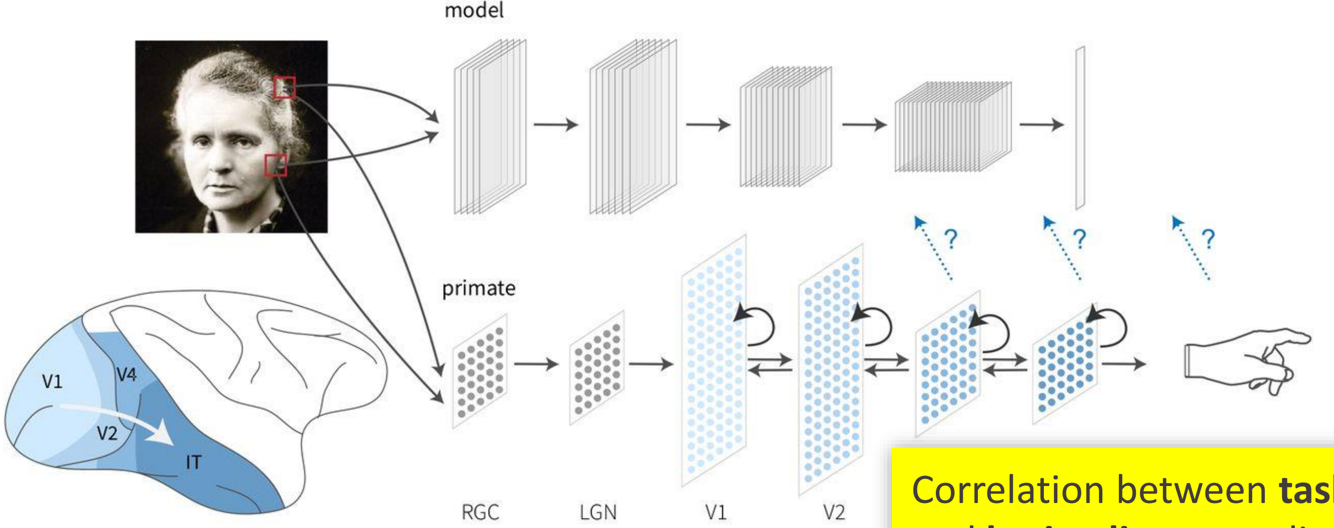
We are far from done!

Not the small-scale circuits

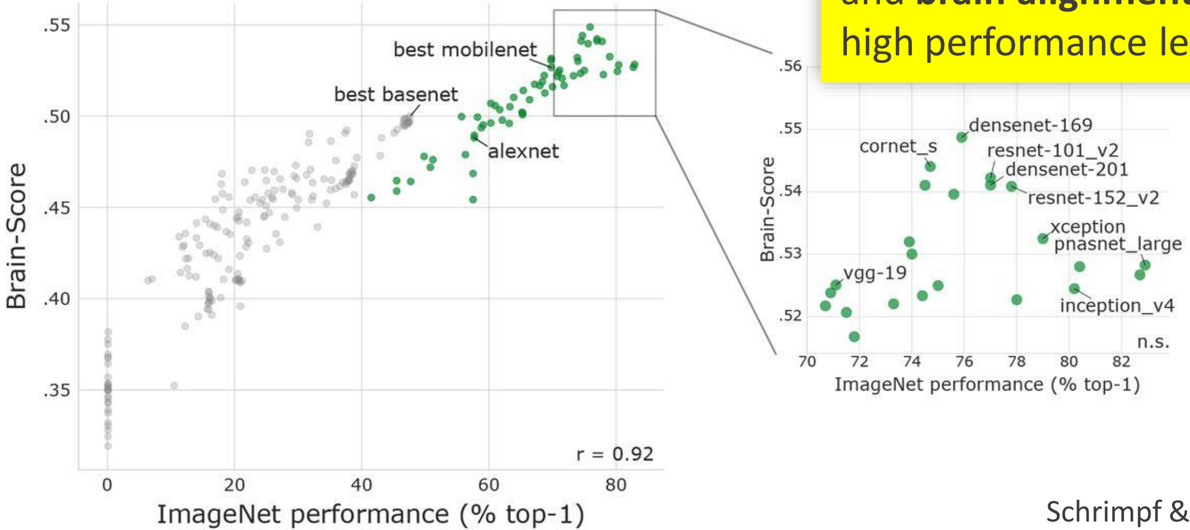
But rather **task performance**, as a combination of **architecture + optimization**

Schrimpf*, Kubilius*, et al. (2018)
Kubilius*, Schrimpf*, et al. (2019)

EPFL



Correlation between **task performance** and **brain alignment** disappears at very high performance levels



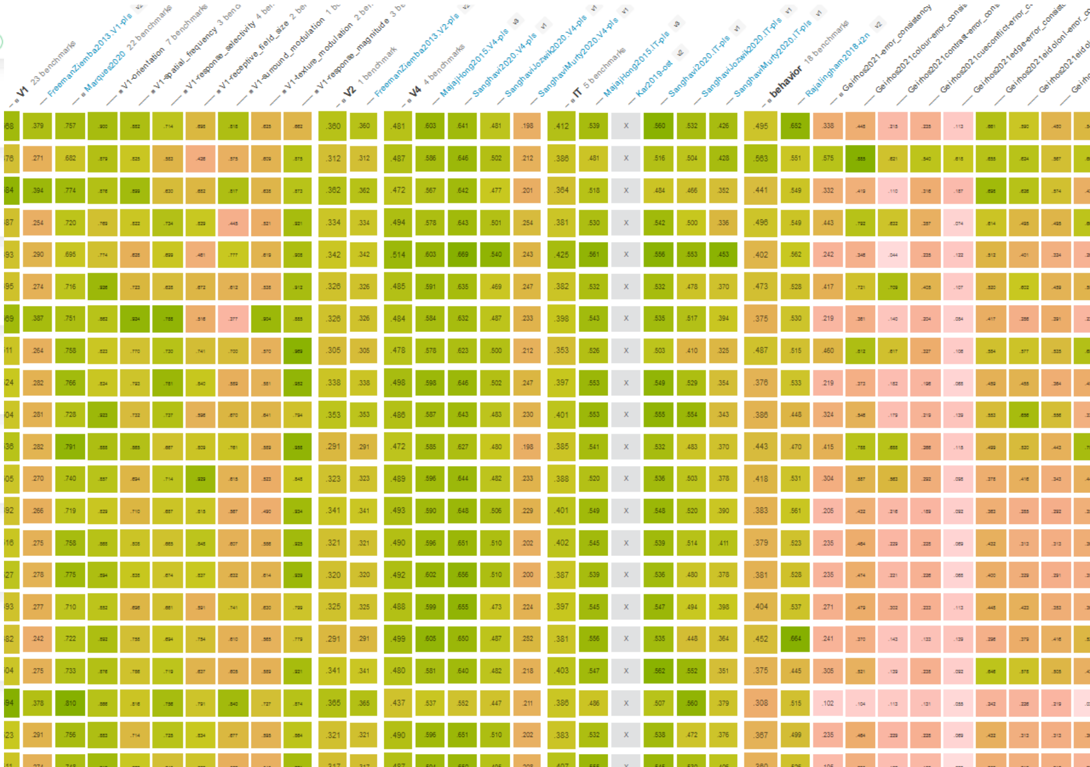
Schrimpf & Kubilius et al. 2018, 2020

Brain-Score 100+ brain & behavior benchmarks, 300+ models

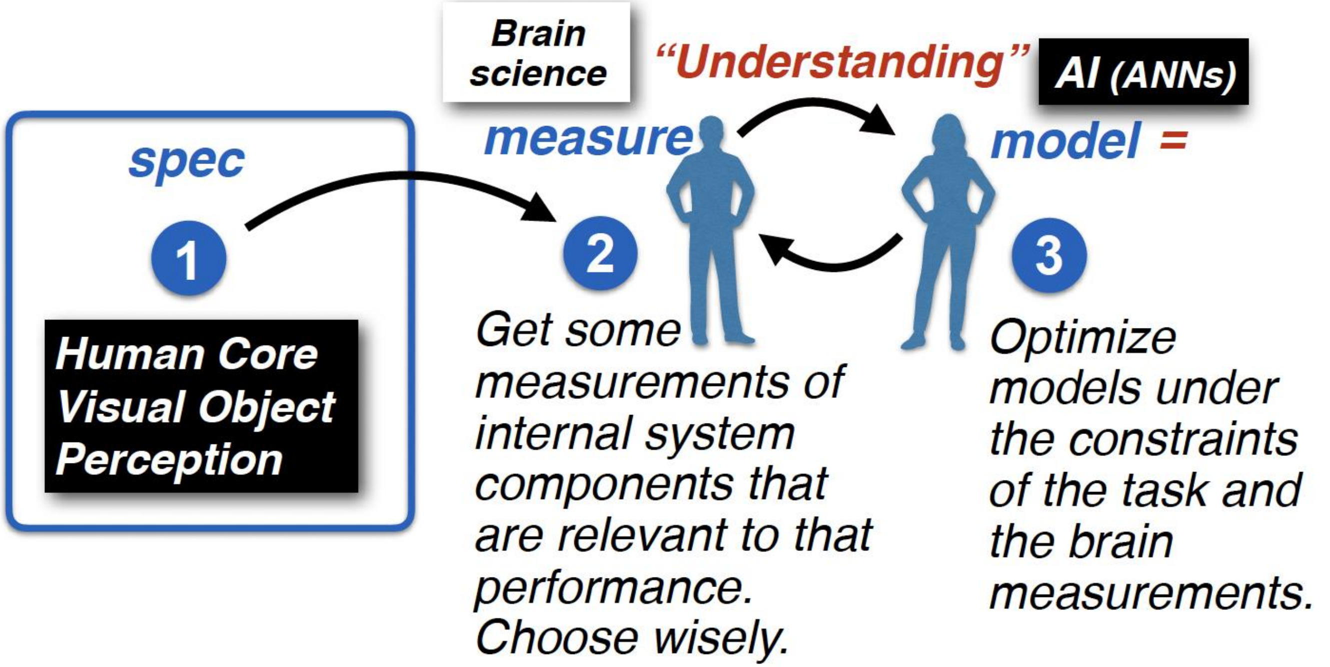
e.g. neural predictions for different image sets, distributional alignments such as spatial frequency, behavioral generalization, …

www.Brain-Score.org

Rank	Model	average 81	neural 38	behavior 43
1	convnext_large_mlp:clip_laio	.467	.378	.555
2	convnext_xlarge:fb_in22k_ft_	.449	.338	.560
3	vit_base_patch16_clip_224:o|	.445	.343	.548
4	vit_large_patch14_clip_224:la	.445	.332	.559
5	vit_large_patch14_clip_224:o	.443	.341	.545
6	vit_base_patch16_clip_224:o|	.442	.352	.532
7	vit_relpos_base_patch16_cls|	.437	.381	.494
8	cvt_cvt-w24-384-in22k_finetuned-in1k_4	.430	.327	.533



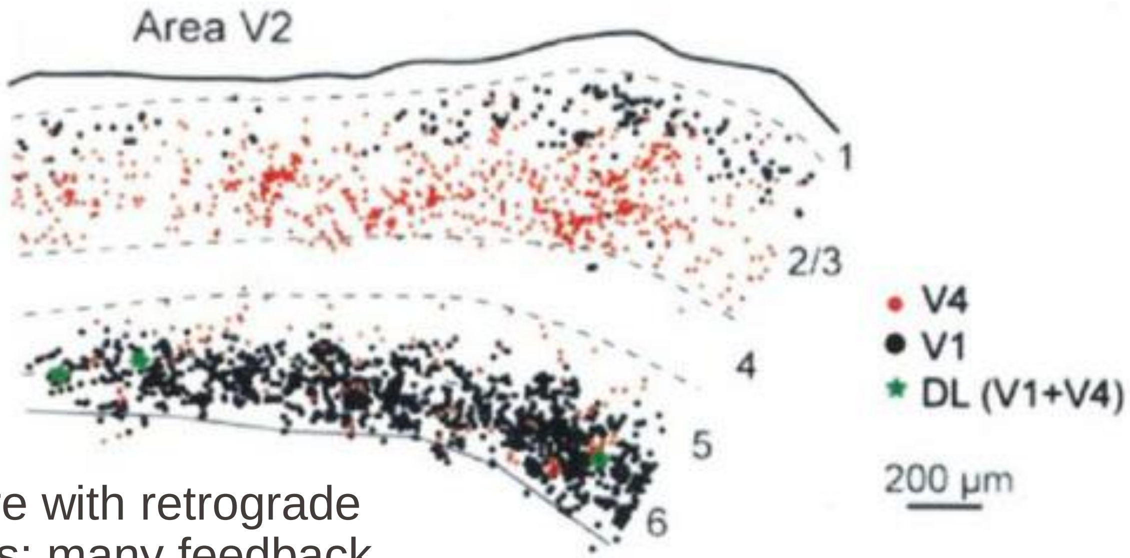
The reverse engineering principle



Slide from Jim DiCarlo, MIT

Feed-forward and recurrent processing

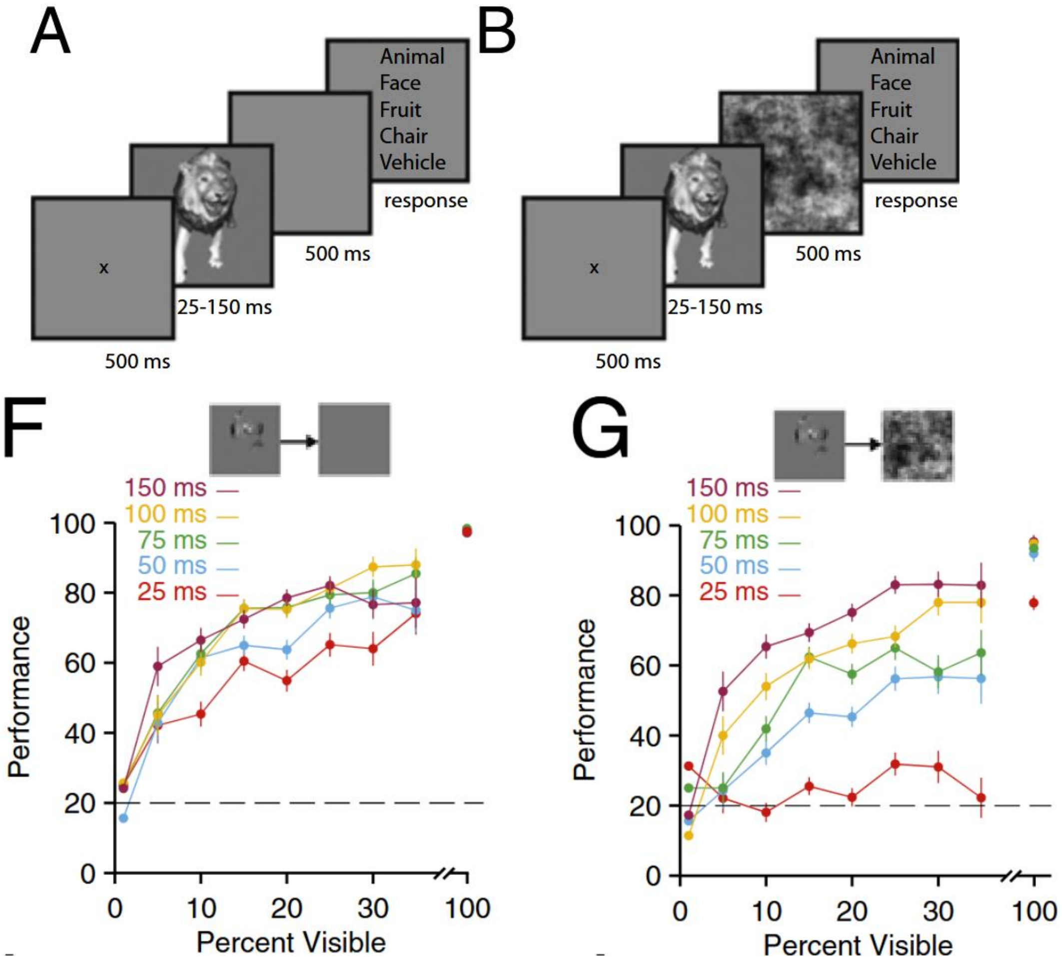
- So far, we have considered the visual system as a hierarchical **feed-forward** feature encoder
- But: many connections in the brain are **recurrent** (depending on the area thought to make up even 80% of all connections)

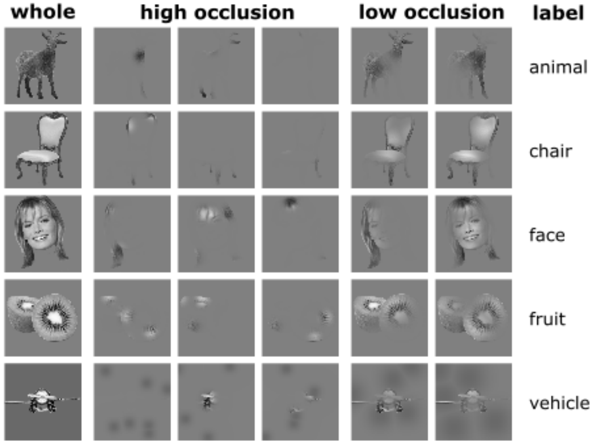


E.g., here with retrograde injections: many feedback connections from V4 onto V2

Rockland 2022

Recurrent processing in the visual system

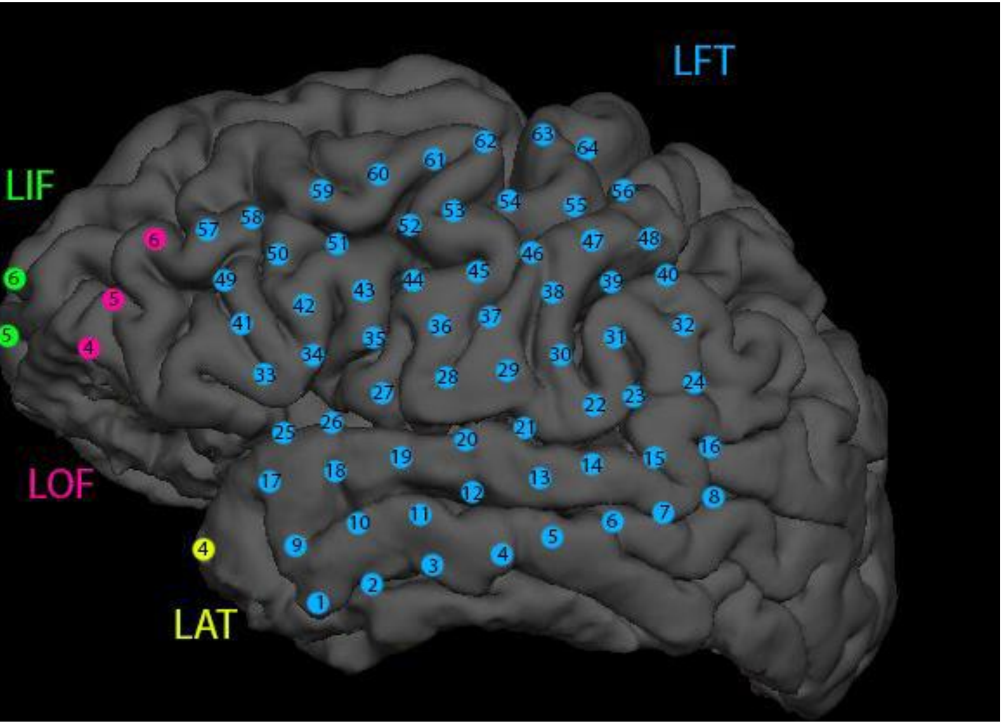




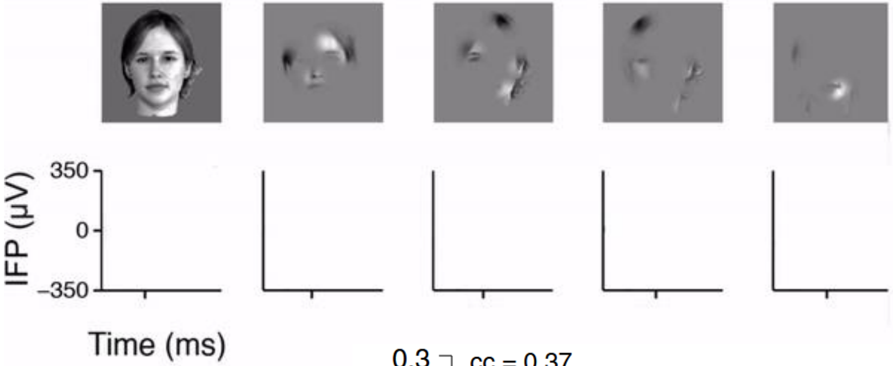
- Humans can classify images better the more visible they are
- Backward mask: thought to disrupt recurrent processing
- With backward masking, performance drops at short presentation times

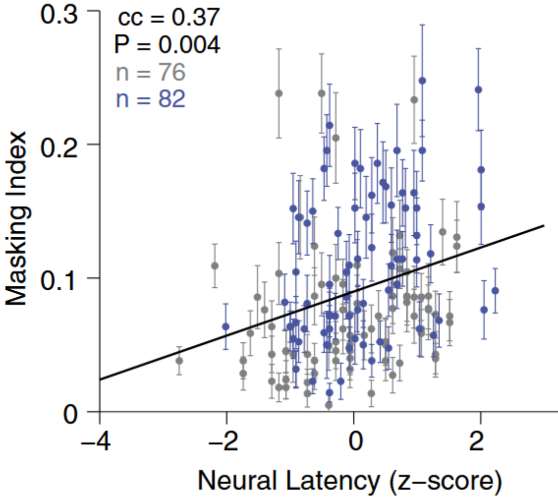
Tang & Schrimpf & Lotter et al. 2018

Recurrent processing in the visual system



Electrode in left fusiform gyrus (face-selective)





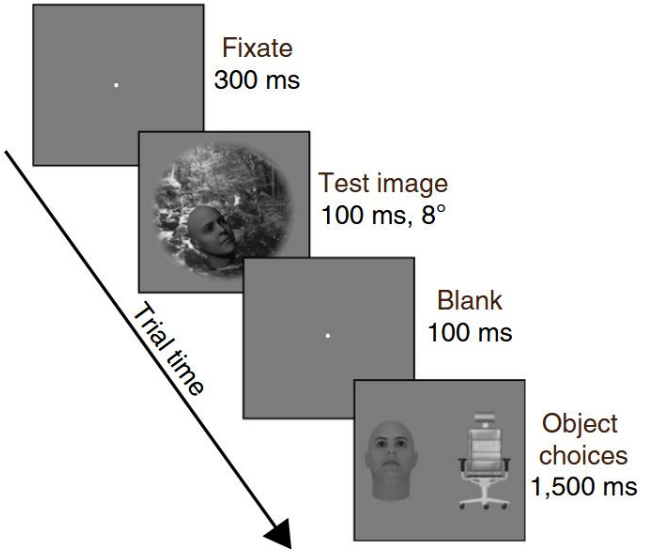
Higher neural latency during occluded image presentation → stronger effect of masking on behavior

Tang & Schrimpf & Lotter et al. 2018

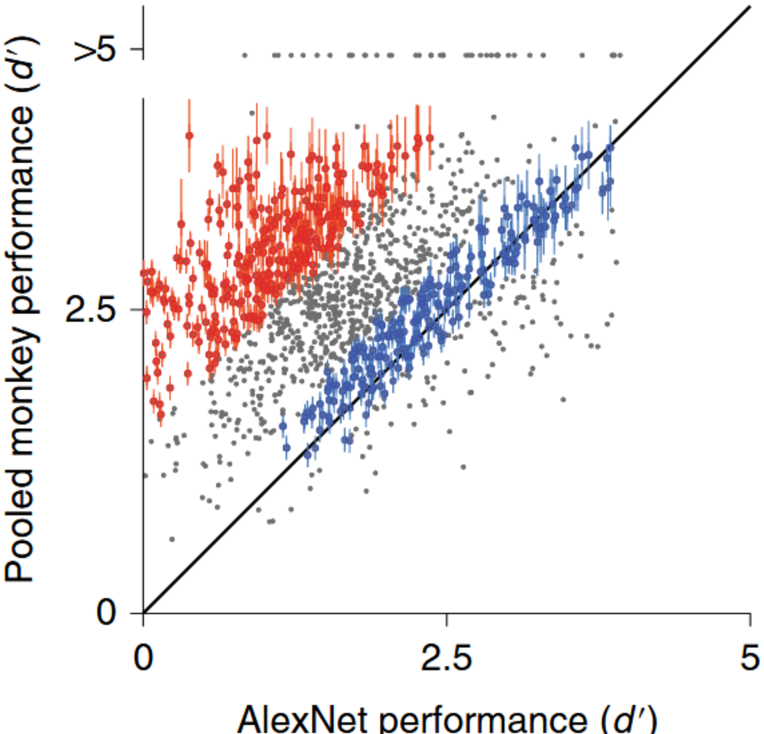
Recurrent processing in primate IT

- Model-guided stimulus selection: choose images where a feed-forward model struggles, but humans perform well.

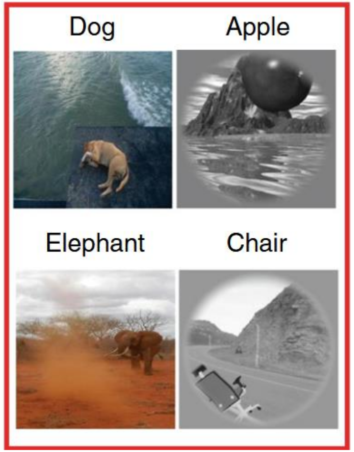
Object discrimination task



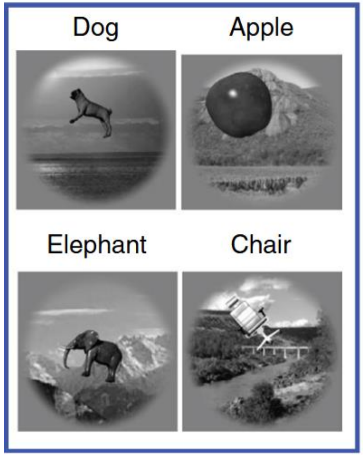
Behavioral comparison
monkeys vs DCNN



Challenge images



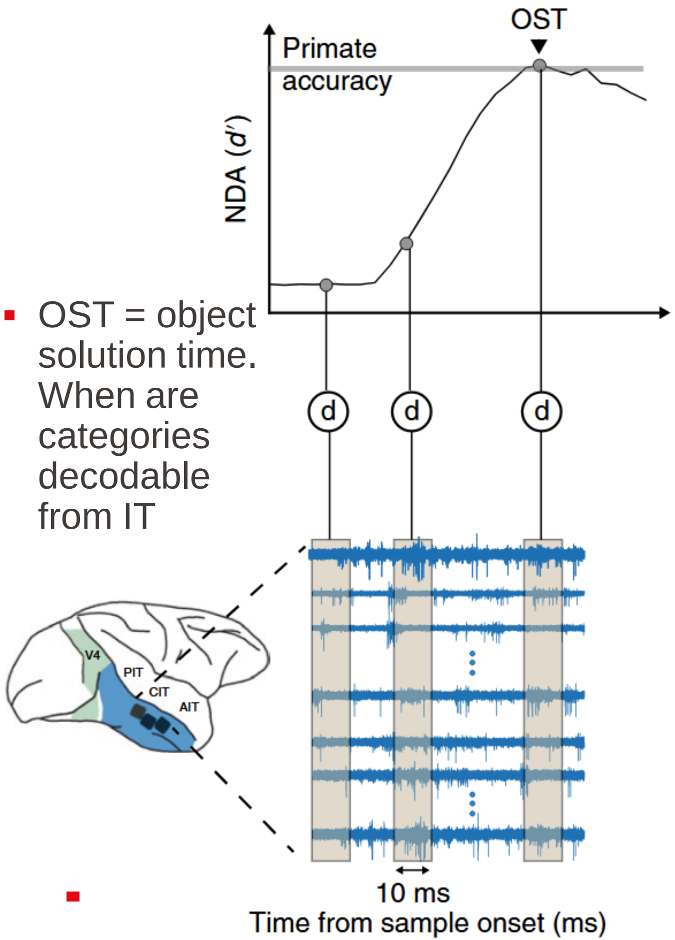
Control images



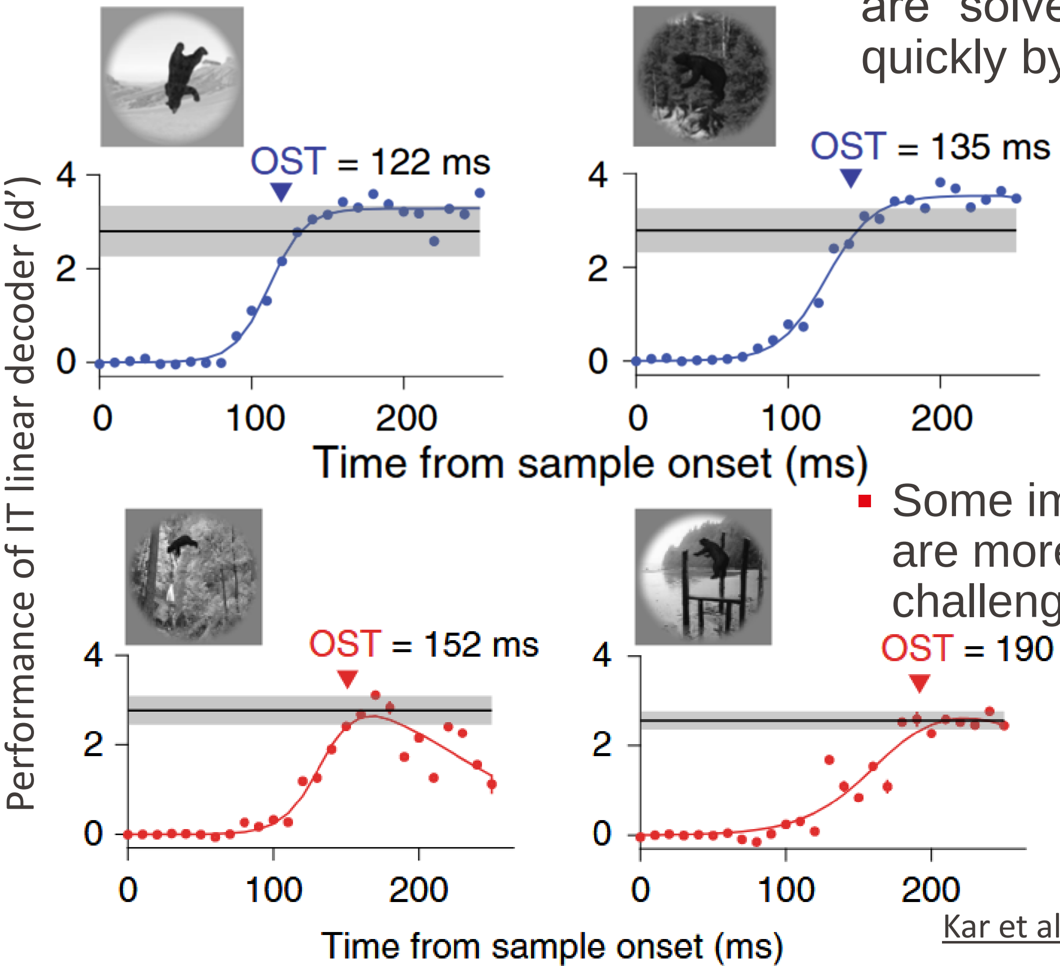
- Kar et al. 2019

Recurrent processing in primate IT

- OST = object solution time. When are categories decodable from IT

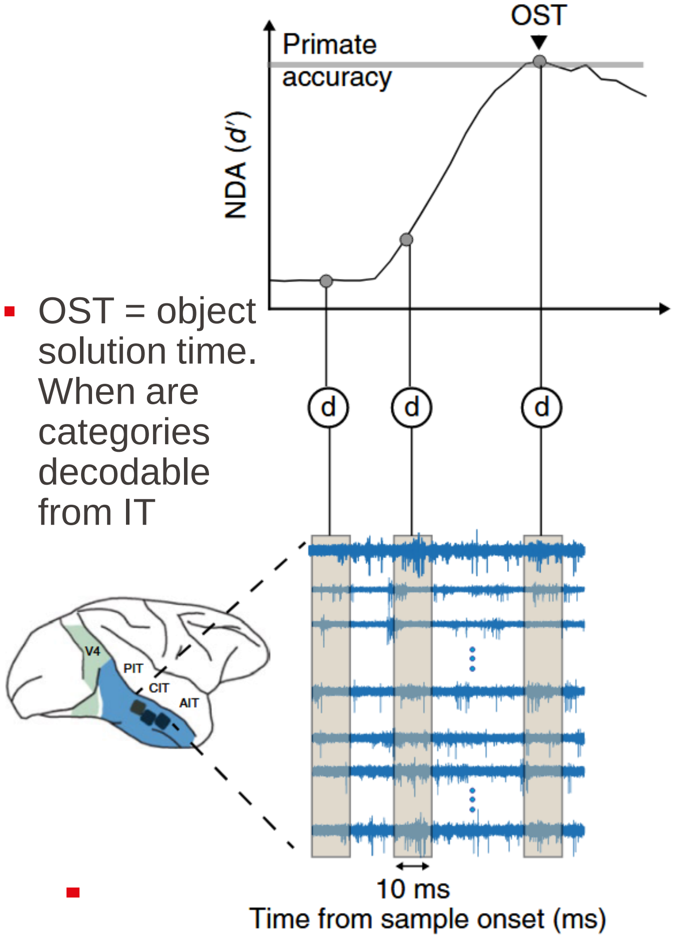


- Some images are "solved" quickly by IT
- Some images are more challenging



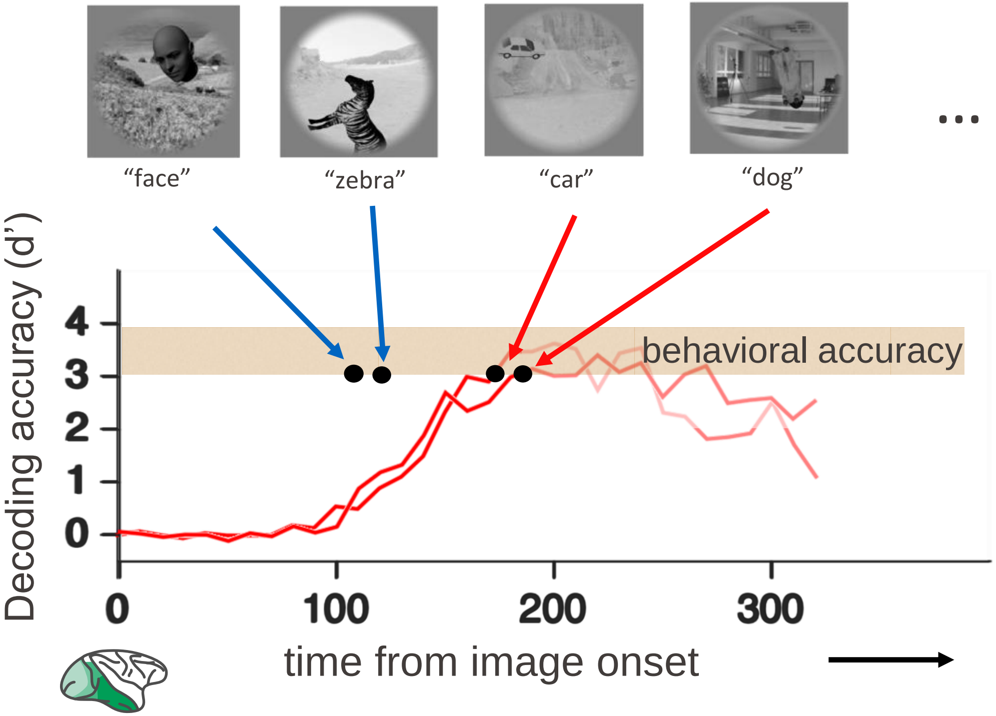
Kar et al. 2019

Recurrent processing in primate IT



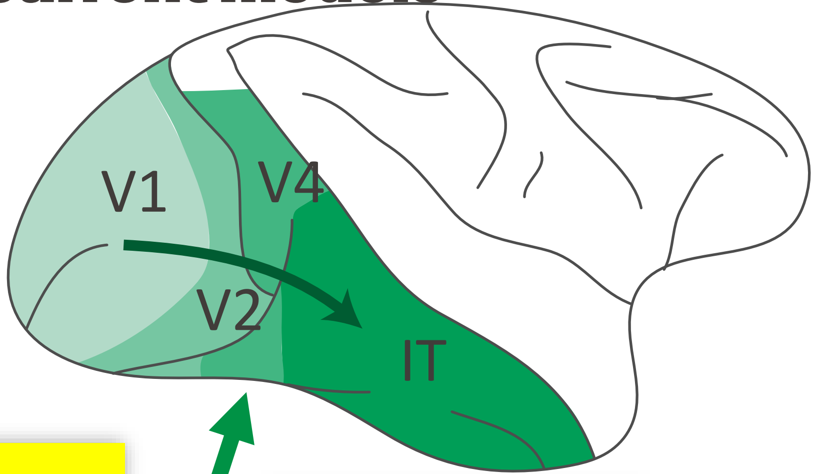
- OST = object solution time. When are categories decodable from IT

- Control images are solved quickly
- Challenge images require more processing



Kar et al. 2019

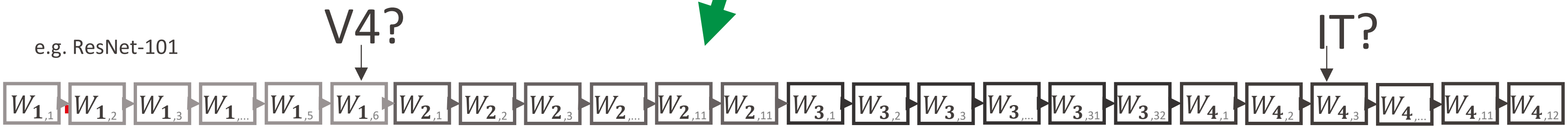
Modeling recurrence: transform feed-forward networks into recurrent models



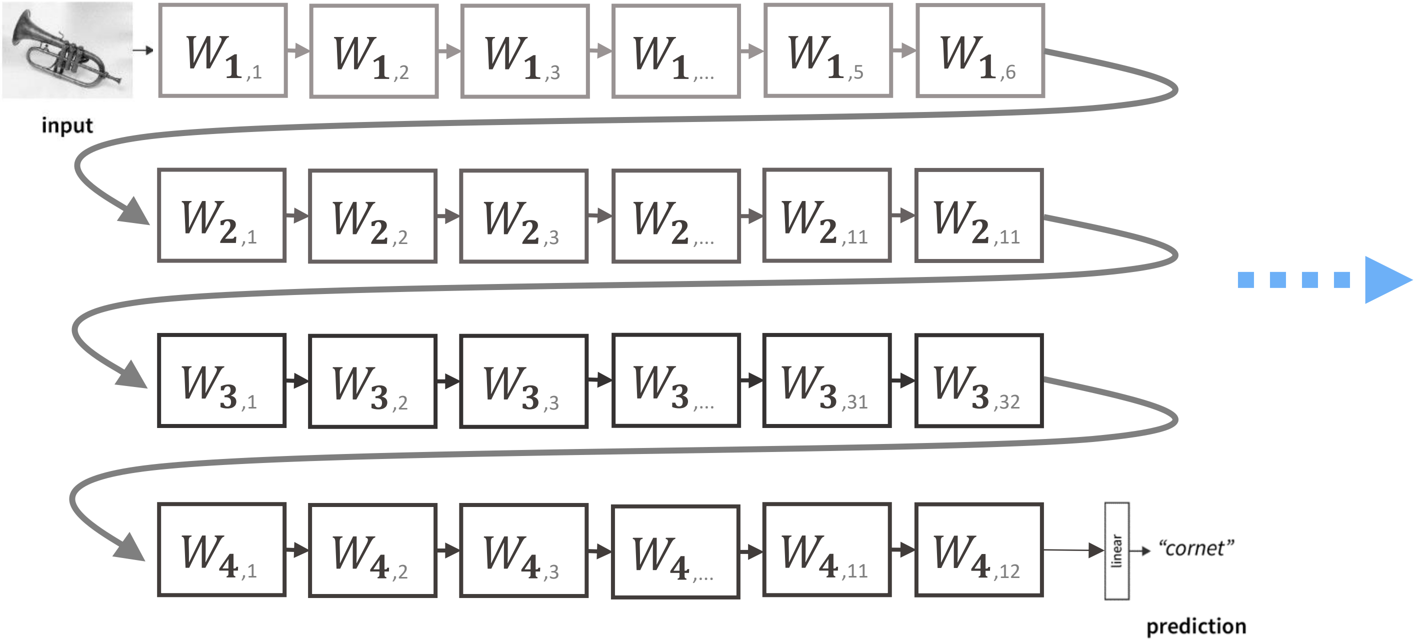
1. difficult to map anatomy

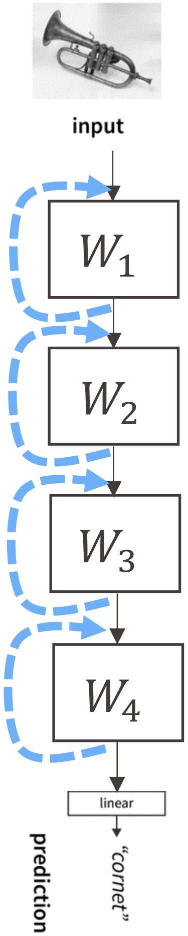
2. no temporal responses

e.g. ResNet-101



Transform feed-forward networks into recurrent models

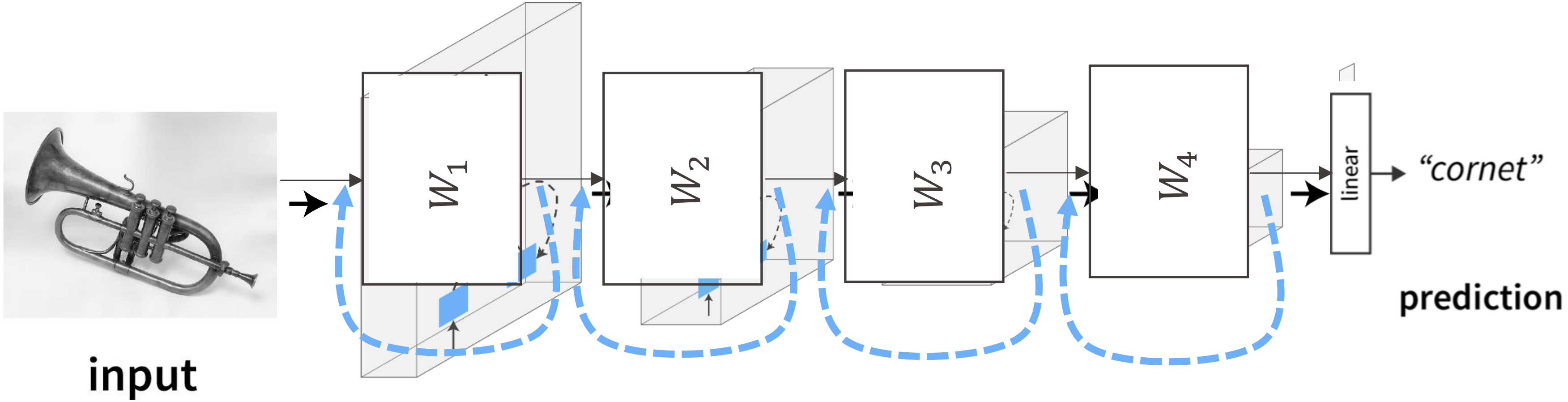




e.g. He, Zhang, Ren, Sun (CVPR 2016)
Huang, Liu, van der Maaten, Weinberger (CVPR 2017)

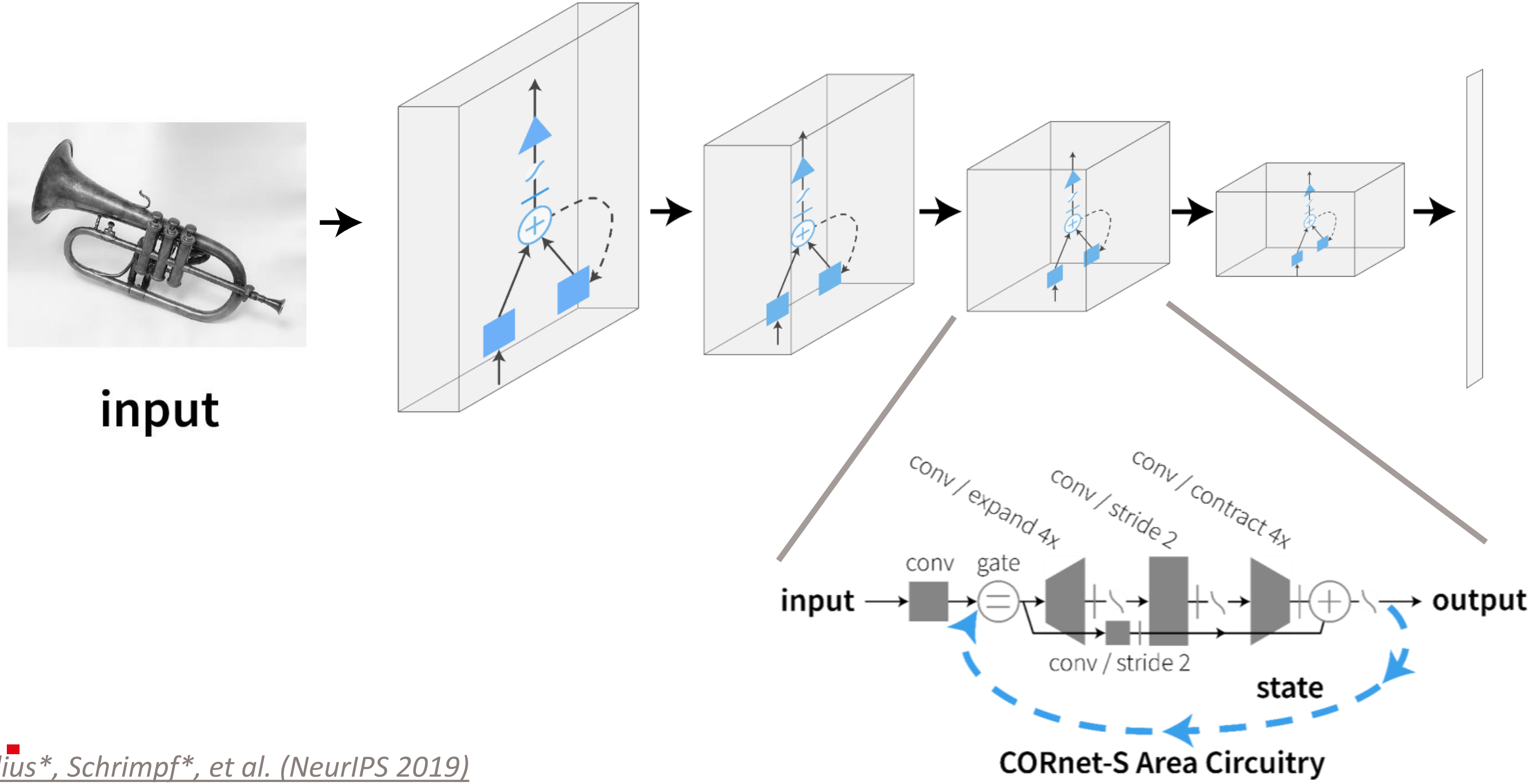
Liang & Hu (CVPR 2015) Liao & Poggio (arXiv 2016) Tang*, Schrimpf*, Lotter* et al. (PNAS 2018) Nayebi*, Bear*, Kubilius* et al. (NIPS 2018)

Modeling recurrence: transform feed-forward networks into recurrent models: CORnet



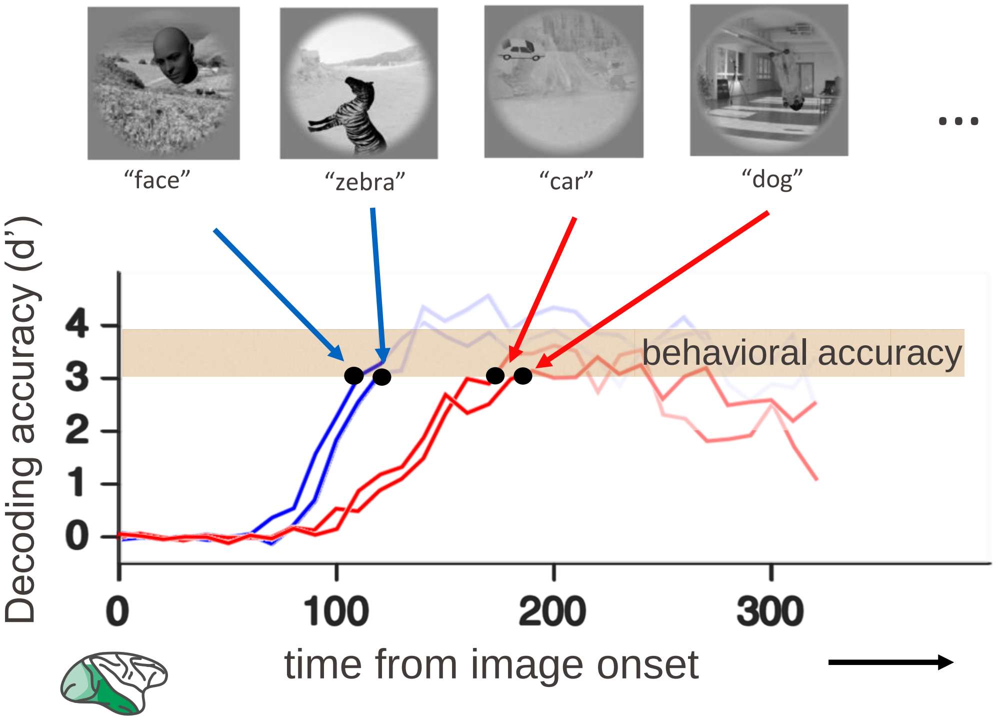
Kubilius*, Schrimpf*, et al. (NeurIPS 2019)

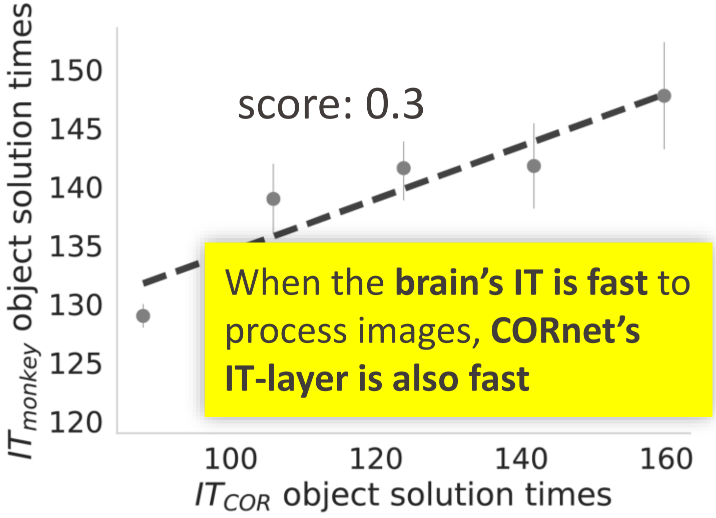
Recurrent CORnet model: compact architecture via recurrence



Kubilius*, Schrimpf*, et al. (NeurIPS 2019)

Recurrent model predicts temporal dynamics in IT





When the **brain's IT is fast** to process images, **CORnet's IT-layer is also fast**

Unlike feedforward models, CORnet-S can **predict neural responses over time**.

Kar, Kubilius, Schmidt, Issa, DiCarlo (Nature Neuroscience 2019)

Kubilius*, Schrimpf*, et al. (NeurIPS 2019)

Take-home messages

- Reverse engineering principle
- Representation learning with goal-driven / task-driven learning provides a normative framework for studying the nervous system.
- Models trained on object-recognition can predict neural activity in ventral pathway during object recognition well. Outperforming previous models by a large margin.
- The work by Yamins, Hong, DiCarlo and collaborators highly influenced vision research and inspired much other work beyond vision.
- Recurrent processing is recruited for more challenging images. Compact recurrent models explain the brain’s temporal dynamics to a first extent.
- In the exercises you’ll recap linear probing & explained variance. Next you’ll start a mini-project modeling IT neurons!