How behavior shapes recurrent circuits across sensory systems and species: from vision to touch

Aran Nayebi

Assistant Professor
Machine Learning Department
Neuroscience Institute (core faculty), Robotics Institute (by courtesy)

NeuroML Workshop

University of Chicago
2026.02.25

ML
MACHINE LEARNING
DEPARTMENT

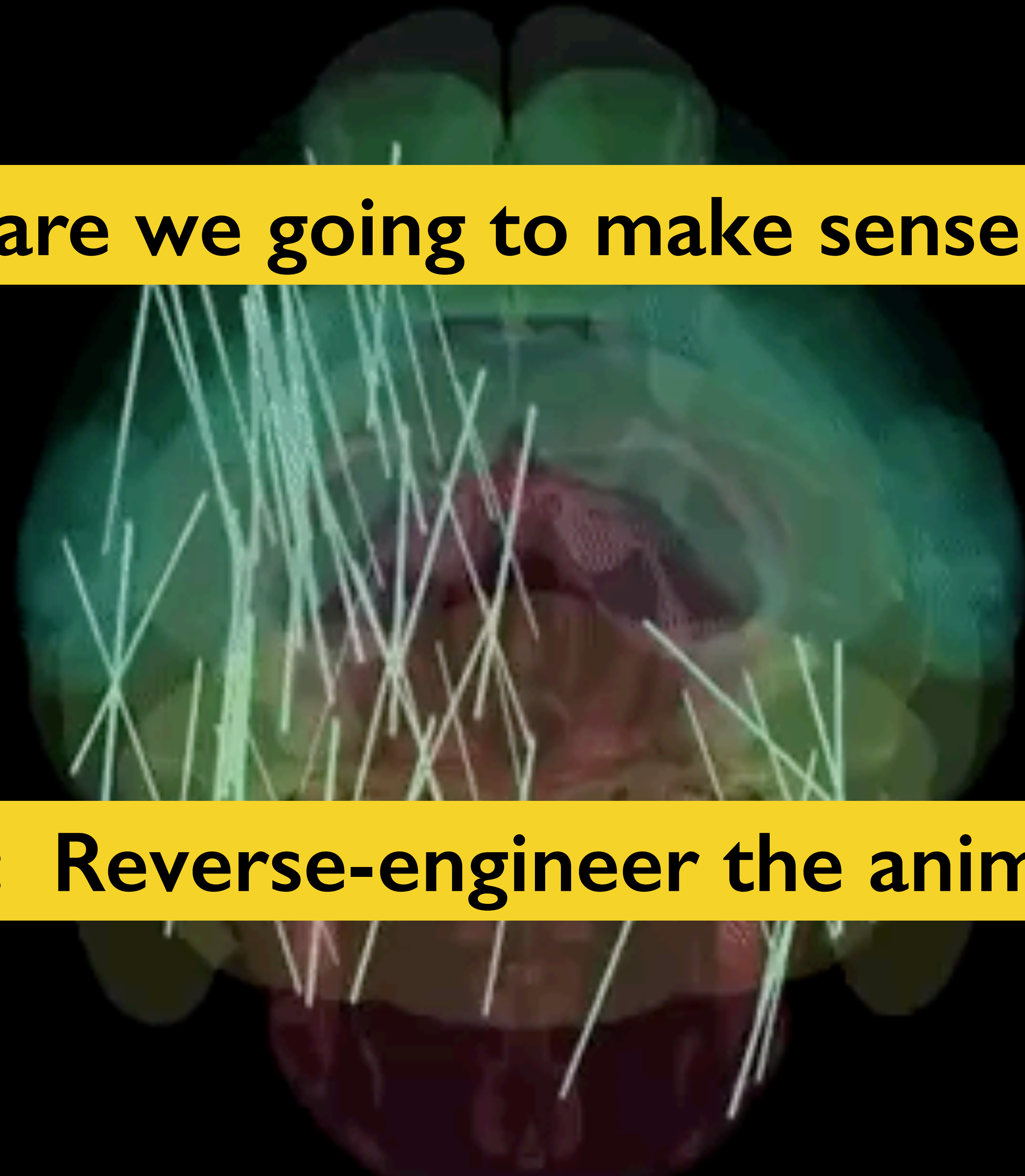
Carnegie Mellon University
Neuroscience Institute

THE ROBOTICS INSTITUTE

Carnegie Mellon
SCHOOL OF COMPUTER SCIENCE

The Brave New World of Large-Scale Neuroscience

Whole brain. . .



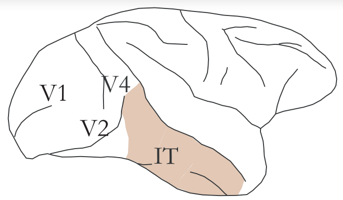
Q: How are we going to make sense of all this data?

A: Reverse-engineer the animal brain.

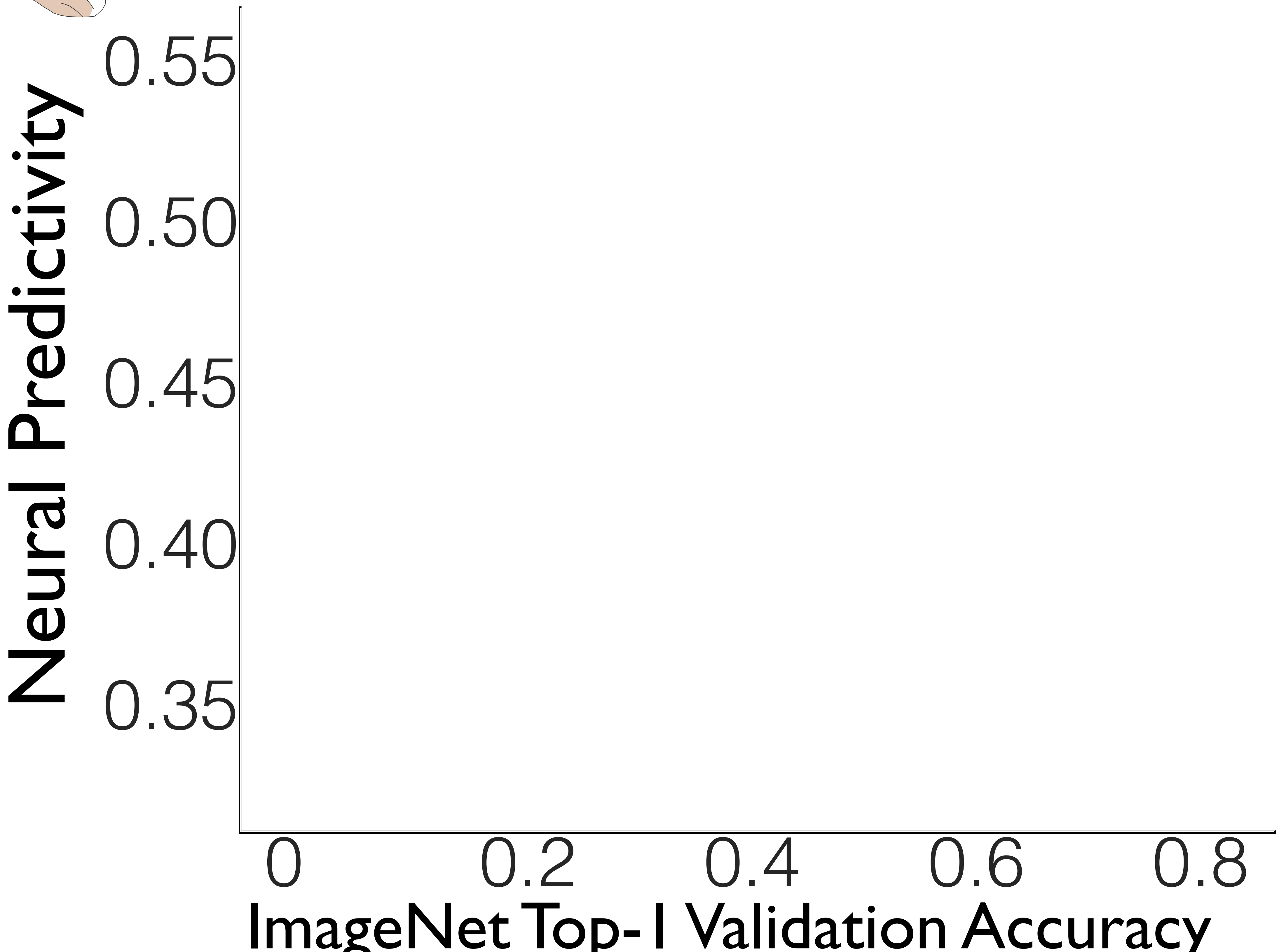
International Brain Laboratory 2022

. . . awake, behaving animals

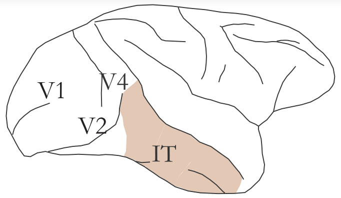
Task Performance Correlated with Neural Predictivity



Schrimpf, Kubilius* et al. 2018*

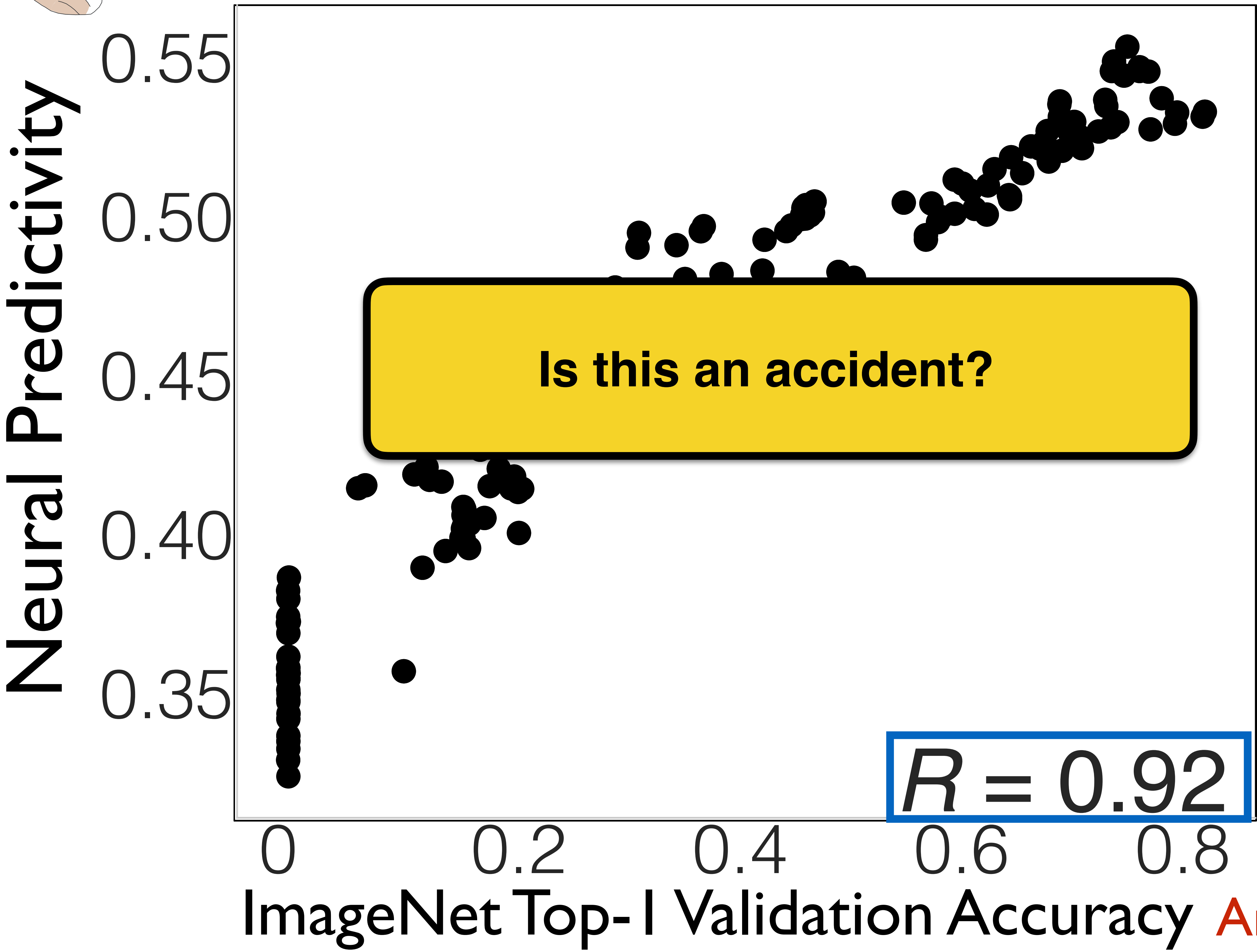


Task Performance Correlated with Neural Predictivity



A Neuroscience Goal

Schrimpf, Kubilius* et al. 2018*



An AI Goal

Task-Optimized Modeling: Four Components

Task-Optimization (ML)

1. $\boxed{\boldsymbol{A} = \textit{architecture class}}$

2. $\boxed{\boldsymbol{T} = \textit{task loss}}$

3. $\boxed{\boldsymbol{D} = \textit{dataset}}$

4. $\boxed{\boldsymbol{L} = \textit{learning rule}}$

Task-Optimized Modeling: Four Components

Task-Optimization (ML)

Neurobiology

1.

$\boldsymbol{A}$ = *architecture class*

2.

$\boldsymbol{T}$ = *task loss*

3.

$\boldsymbol{D}$ = *dataset*

4.

$\boldsymbol{L}$ = *learning rule*

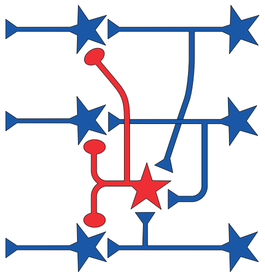
Task-Optimized Modeling: Four Components

Task-Optimization (ML)

Neurobiology

1.

A = *architecture class* **= *circuit neuroanatomy***



2.

T = *task loss*

3.

D = *dataset*

4.

L = *learning rule*

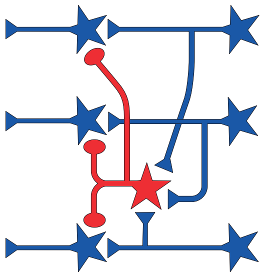
Task-Optimized Modeling: Four Components

Task-Optimization (ML)

Neurobiology

1.

A = *architecture class* ***= circuit neuroanatomy***



2.

T = *task loss* ***= ecological niche/behavior***

3.

D = *dataset*

4.

L = *learning rule*

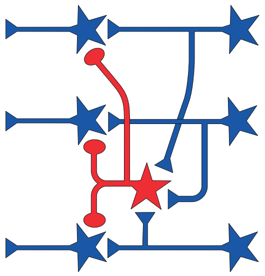
Task-Optimized Modeling: Four Components

Task-Optimization (ML)

Neurobiology

1.

A = *architecture class* ***= circuit neuroanatomy***

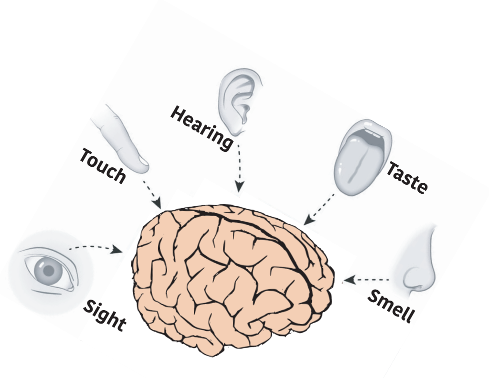


2.

T = *task loss* ***= ecological niche/behavior***

3.

D = *dataset* ***= environment***



4.

L = *learning rule*

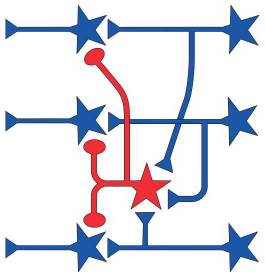
Task-Optimized Modeling: Four Components

<u>Task-Optimization (ML)</u>

<u>Neurobiology</u>

1.

A = *architecture class* **= *circuit neuroanatomy***

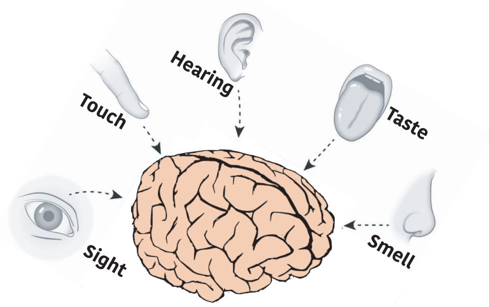


2.

T = *task loss* **= *ecological niche/behavior***

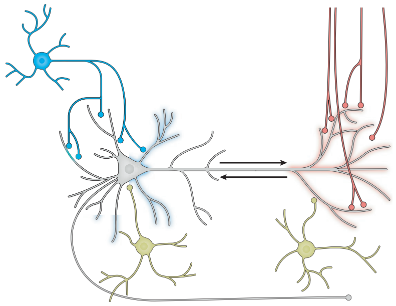
3.

D = *dataset* **= *environment***



4.

L = *learning rule* **= *natural selection + synaptic plasticity***



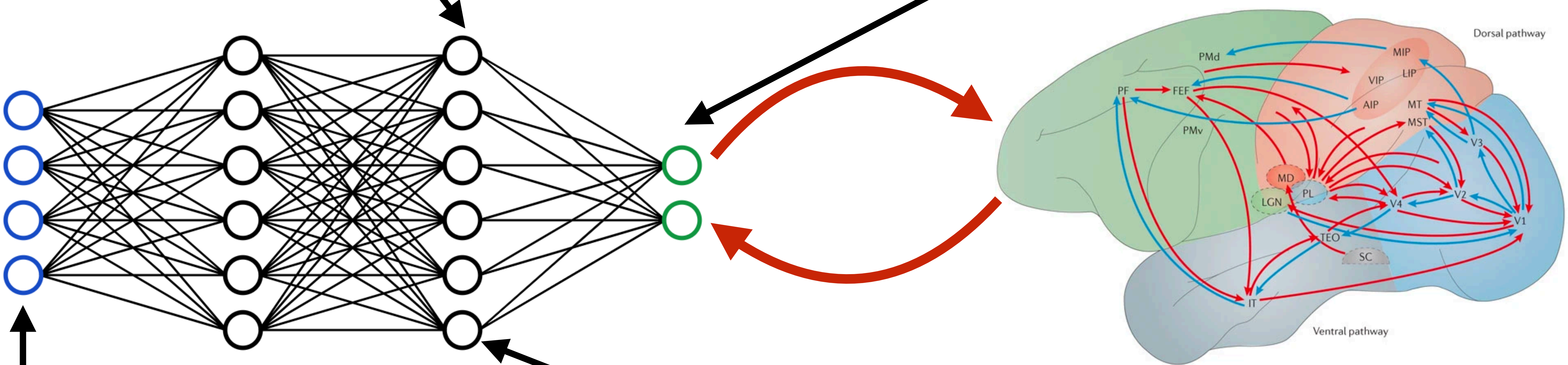
Task-Optimized Modeling: Four Components

L = *learning rule*

"Natural selection + plasticity"

T = *task loss*

"Ecological niche/ behavior"



"Environment"

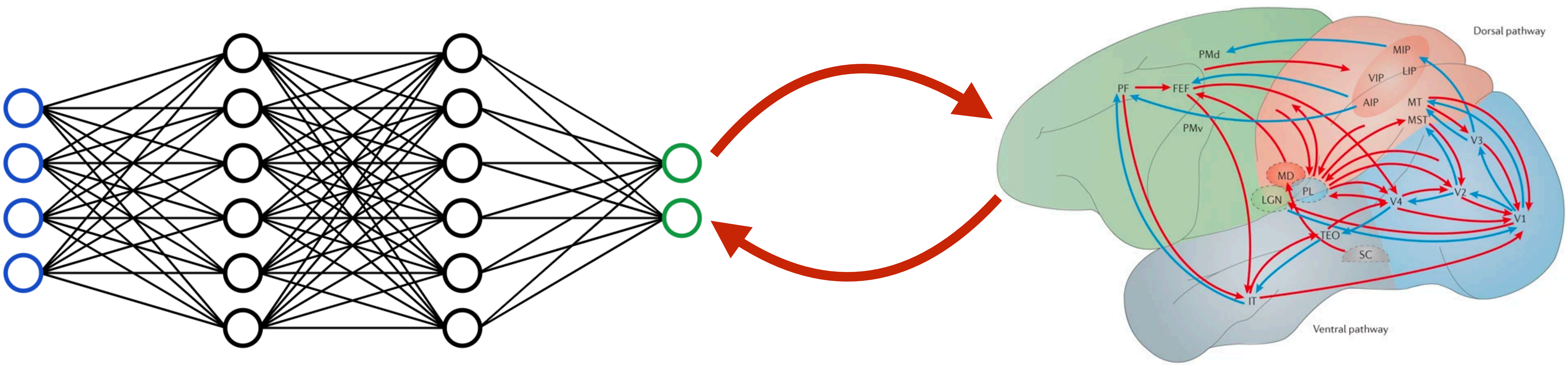
D = *data stream*

"Circuit"

A = *architecture class*

Task-Optimized Modeling

Design ML Algorithms Optimized to Perform Organism's Behavior under Organism's Constraints



Yields:

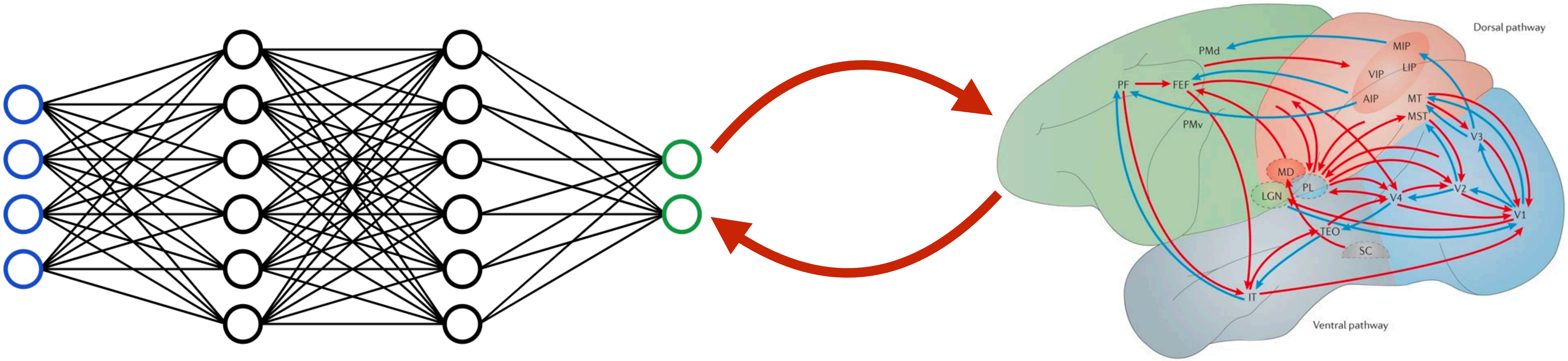
Quantitatively Accurate & Practically Useful Brain Models

AND

Principles of *Why* Neural Responses Are As They Are

Task-Optimized Modeling

Design ML Algorithms Optimized to Perform Organism's Behavior under Organism's Constraints



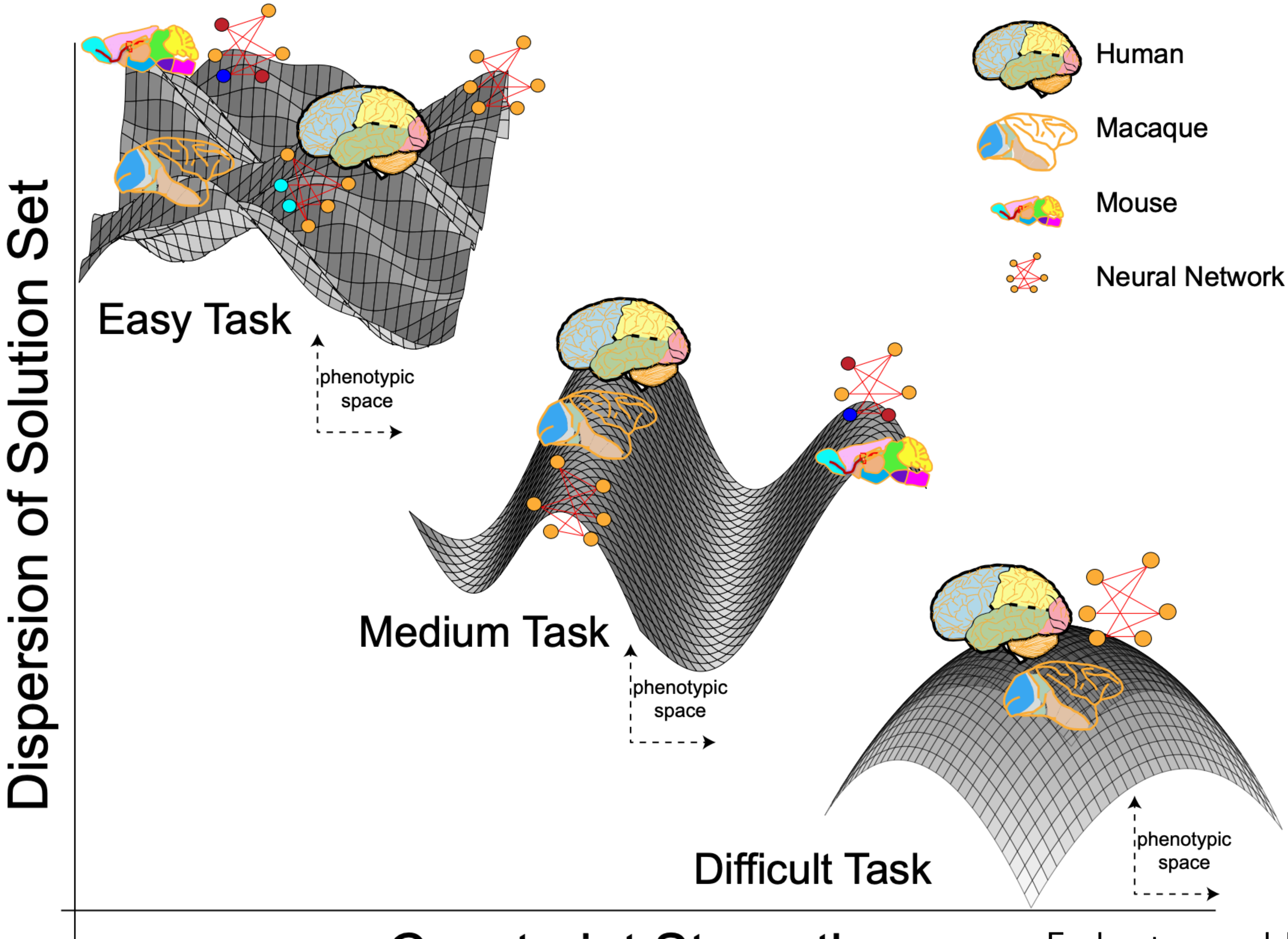
Yields:

Quantitatively Accurate & Practically Useful Brain Models

AND

Principles of *Why* Neural Responses Are As They Are

Contravariance Principle: The Harder the Task, the Less Solutions!



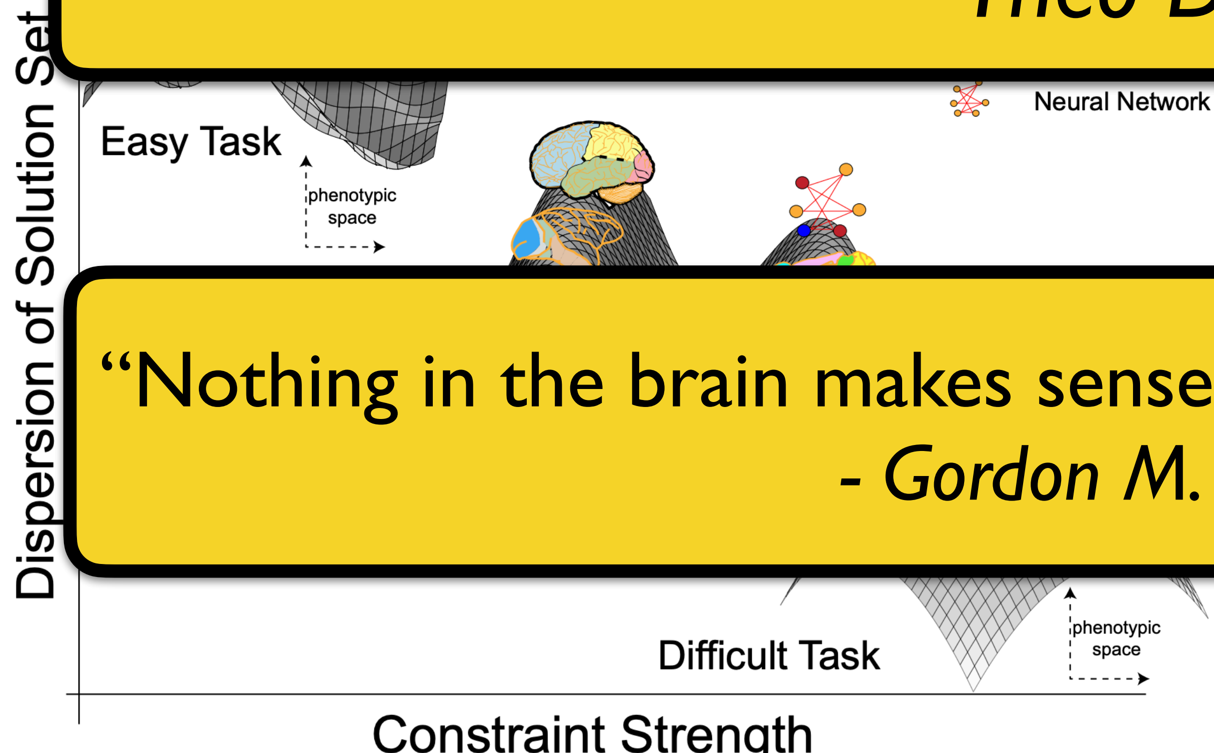
Rosa Cao

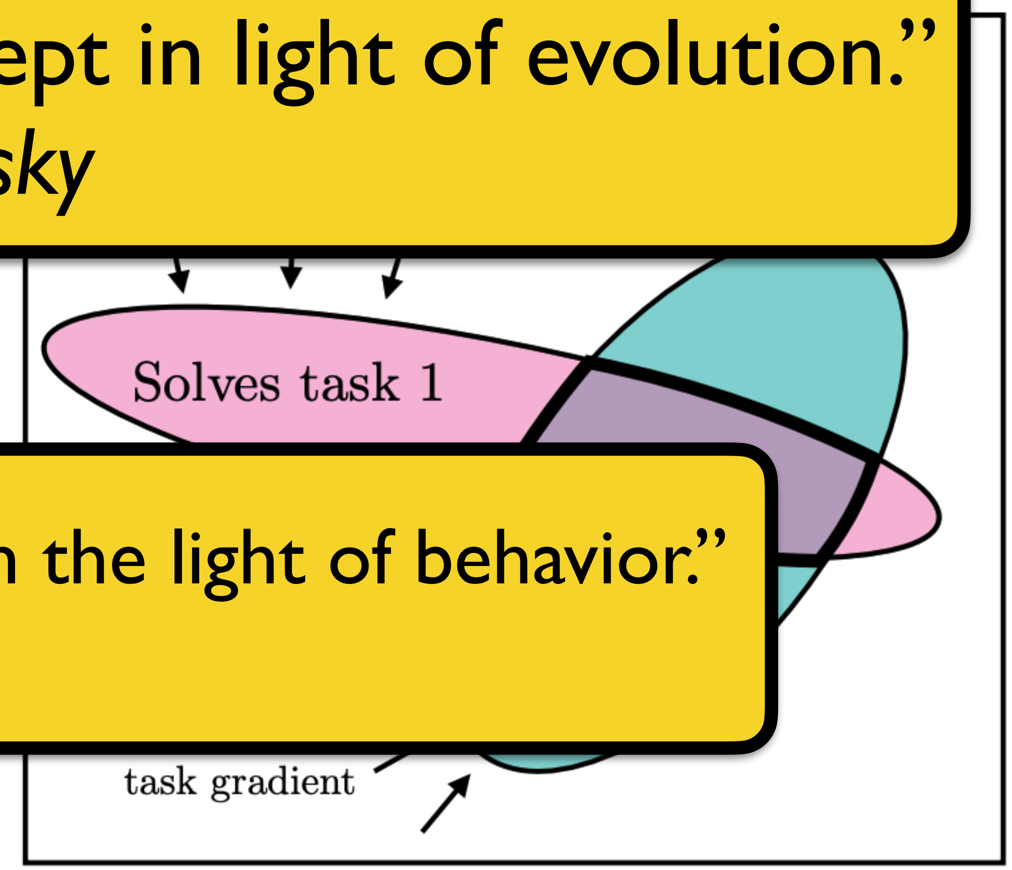
Daniel Yamins

Explanatory models in neuroscience:
Part 2 – Constraint-based intelligibility

Platonic Representation Hypothesis is the AI version of Contravariance

“Nothing in biology makes sense except in light of evolution.”
- Theo Dobzhansky





“Nothing in the brain makes sense except in the light of behavior.”
- Gordon M. Shepherd

Figure 6. **The Multitask Scaling Hypothesis:** Models trained ... ted to pressure to ... s.

<u>Our (slightly) modified credo:</u>
“Nothing in (computational) neuroscience makes sense except in light of task-optimization.”

The Platonic Representation Hypothesis

Minyoung Huh [*1] Brian Cheung [*1] Tongzhou Wang [*1] Phillip Isola [*1]

The Multitask Scaling Hypothesis

There are fewer representations that are competent for N tasks than there are for $M < N$ tasks. As we train more general models that solve more tasks at once, we should expect fewer possible solutions.

Outline

- Role of Recurrent Processing During Object Recognition
- Recurrent Processing Best Explains Tactile Perception

Outline

- Role of Recurrent Processing During Object Recognition
- Recurrent Processing Best Explains Tactile Perception

Role of Recurrent Processing During Object Recognition

A. Nayebi*, D. Bear*, J. Kubilius*, *et al.*
Task-Driven Convolutional Recurrent Models of the
Visual System. *NeurIPS 2018*

A. Nayebi, *et al.*
Recurrent Connections in the Primate Ventral Visual Stream Mediate a Tradeoff
Between Task Performance and Network Size During Core Object Recognition.
Neural Computation 2022

Daniel Yamins

Daniel Bear

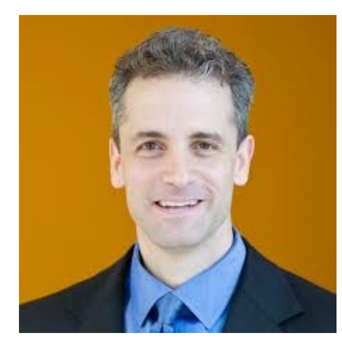
Jonas Kubilius

Kohitij Kar

Surya Ganguli

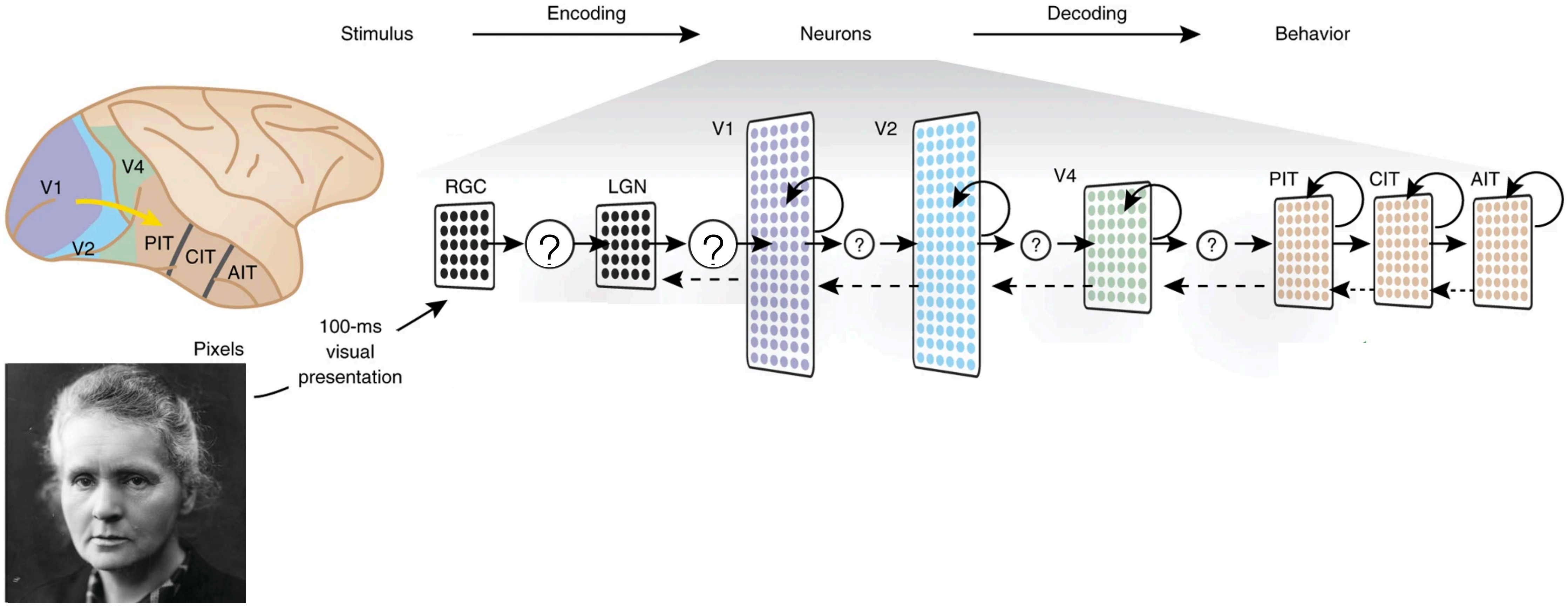
Javier Sagastuy

David Sussillo

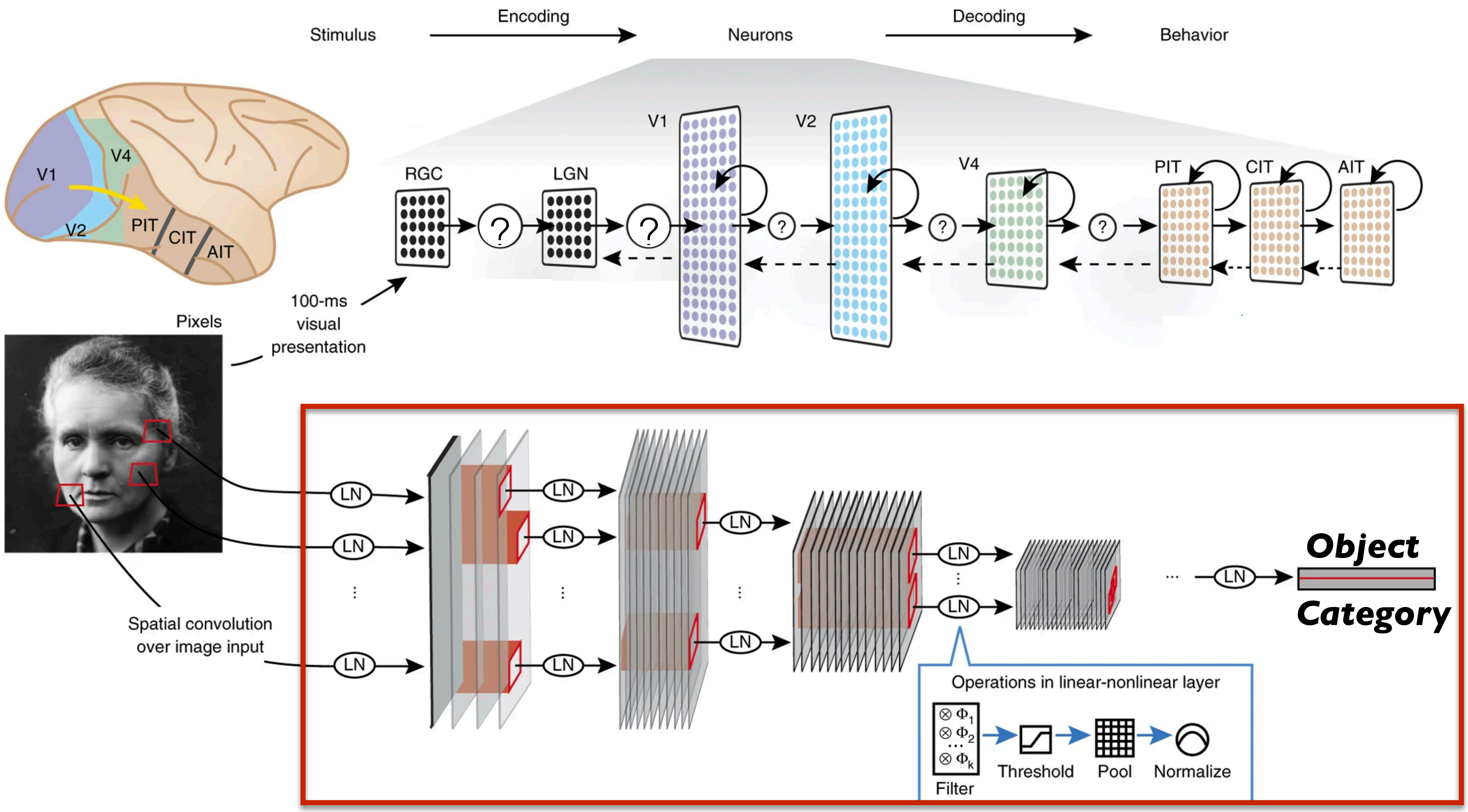


Jim DiCarlo

Primate Ventral Stream Implements Object Recognition



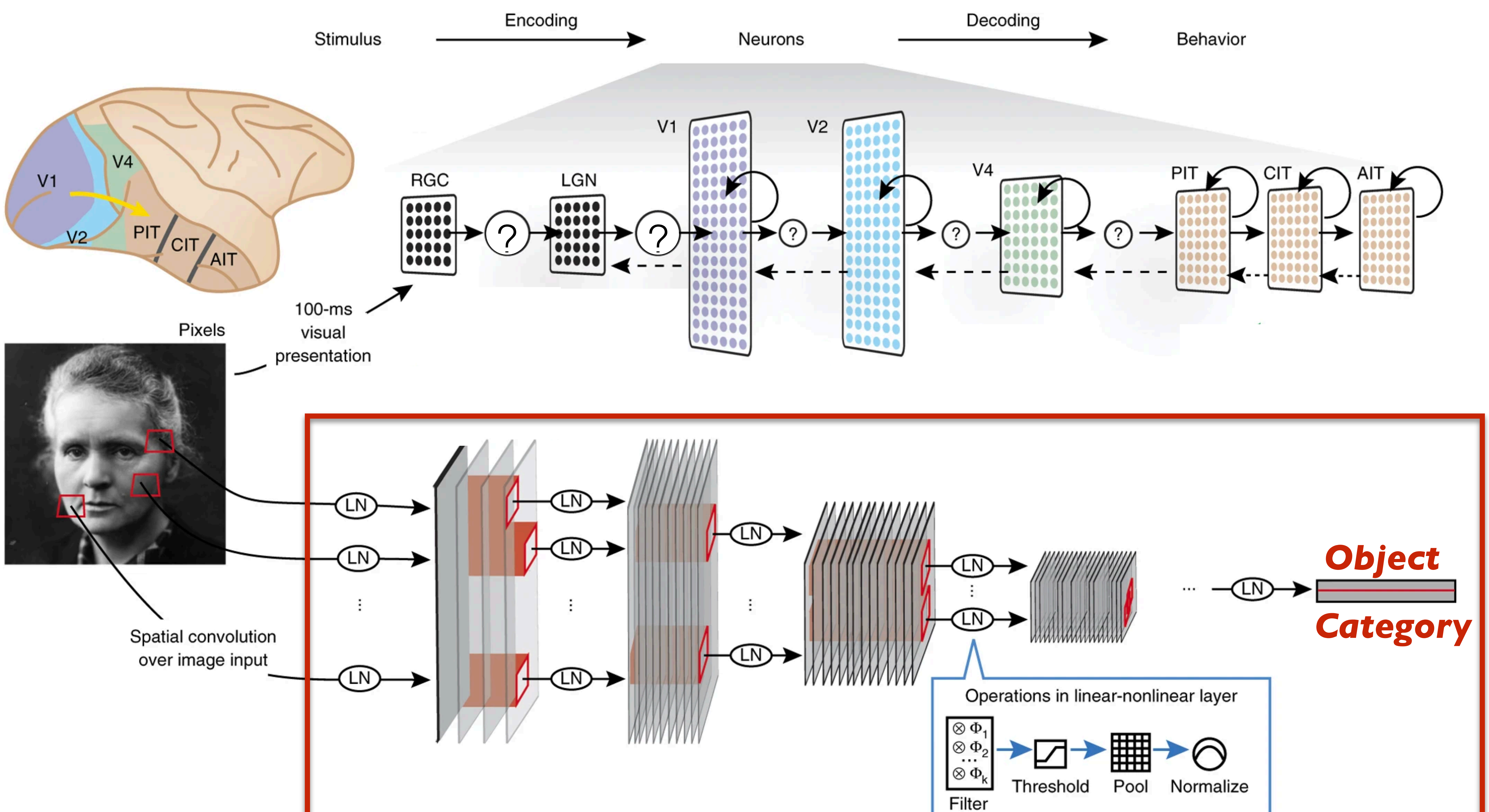
CNNs as Models of Primate Object Recognition



CNNs are inspired by visual neuroscience:

1) **hierarchy**
2) **retinotopy** (spatially tiled)

CNNs as Models of Primate Object Recognition



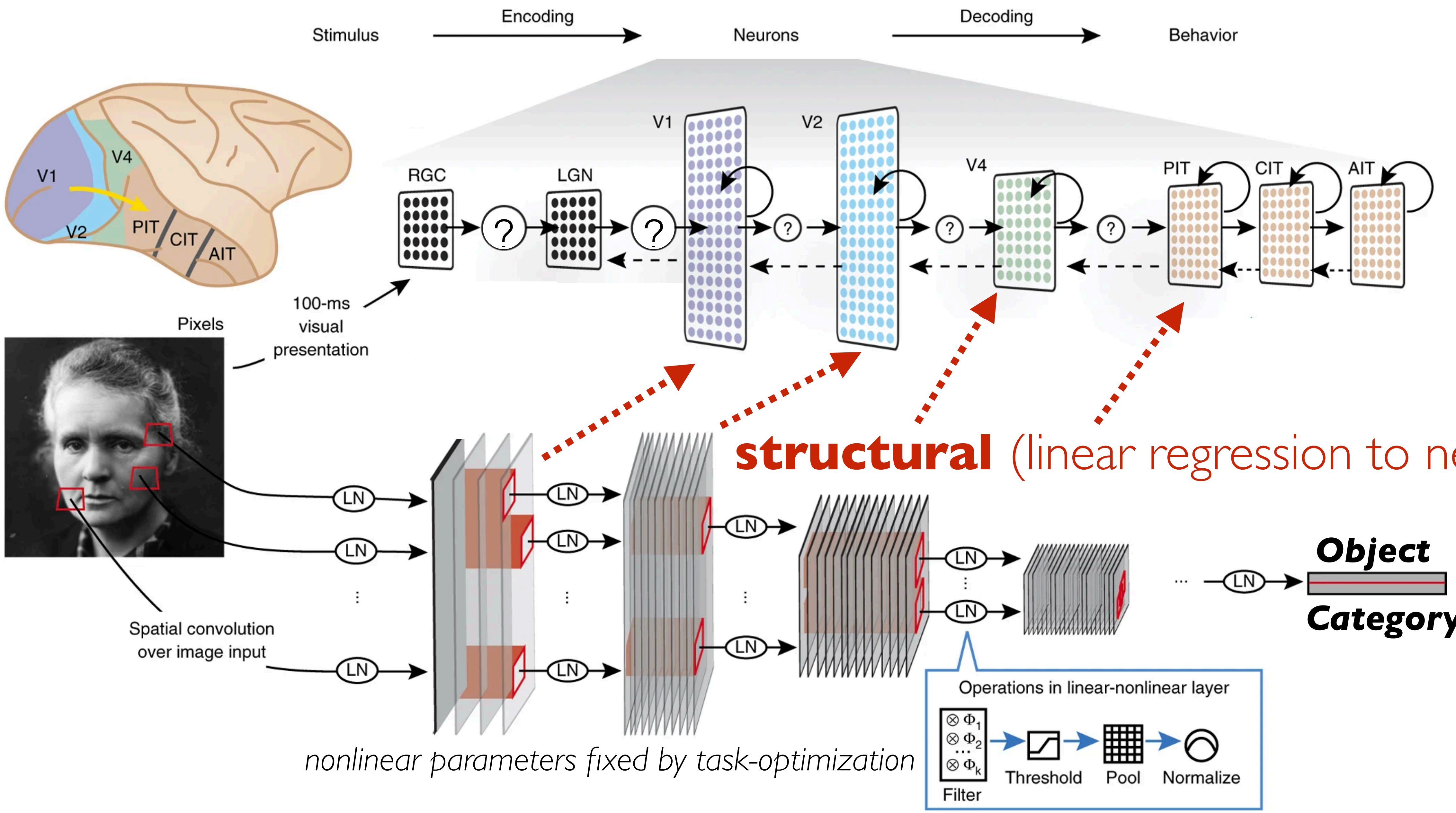
CNNs are inspired by visual neuroscience:

1) **hierarchy**

2) **retinotopy** (spatially tiled)

functional (performs behavior)

CNNs as Models of Primate Object Recognition



structural (linear regression to neurons)

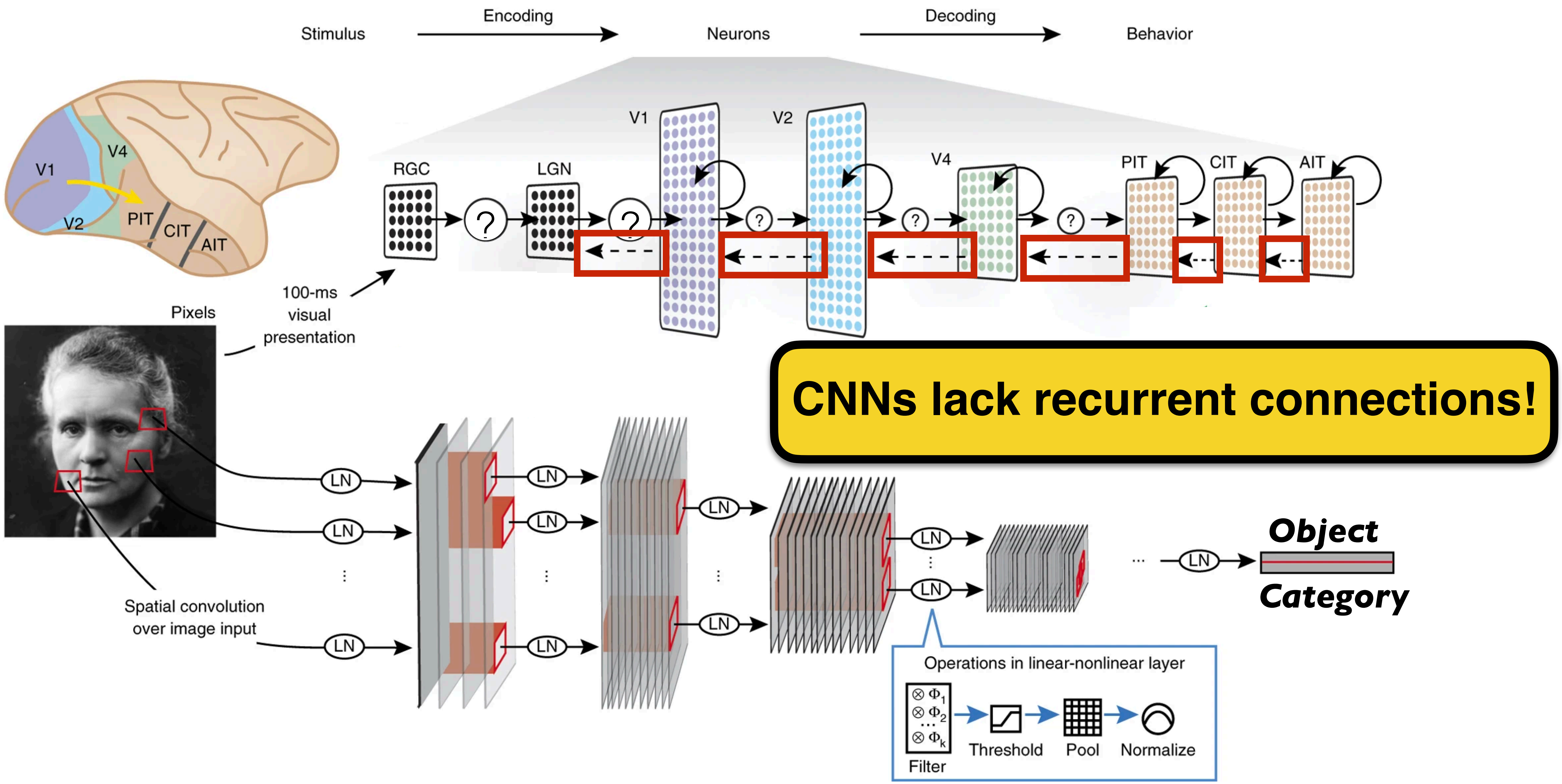
nonlinear parameters fixed by task-optimization

CNNs are inspired by visual neuroscience:

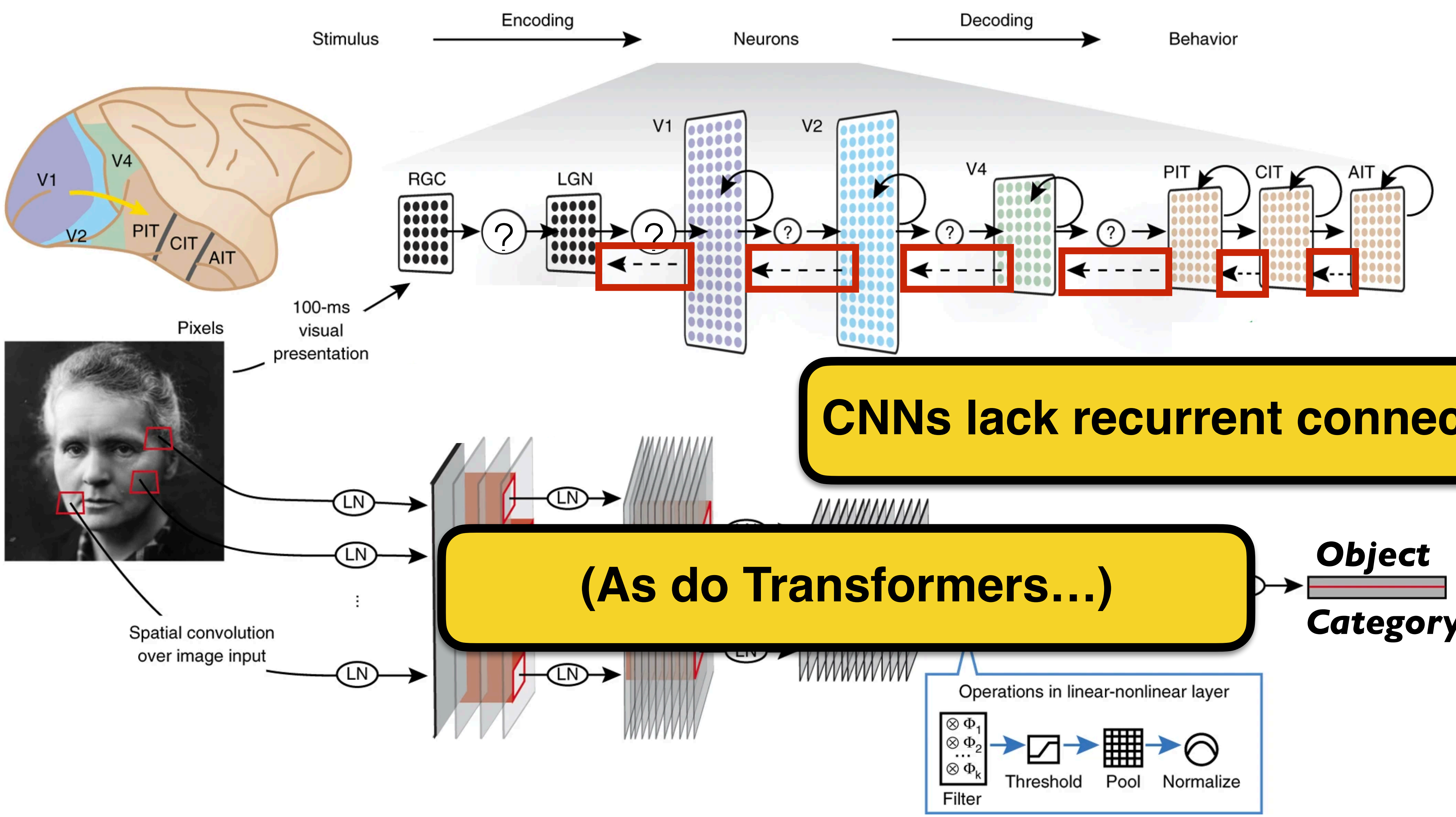
1) **hierarchy**
2) **retinotopy** (spatially tiled)

functional (performs behavior)

CNNs as Models of Primate Object Recognition



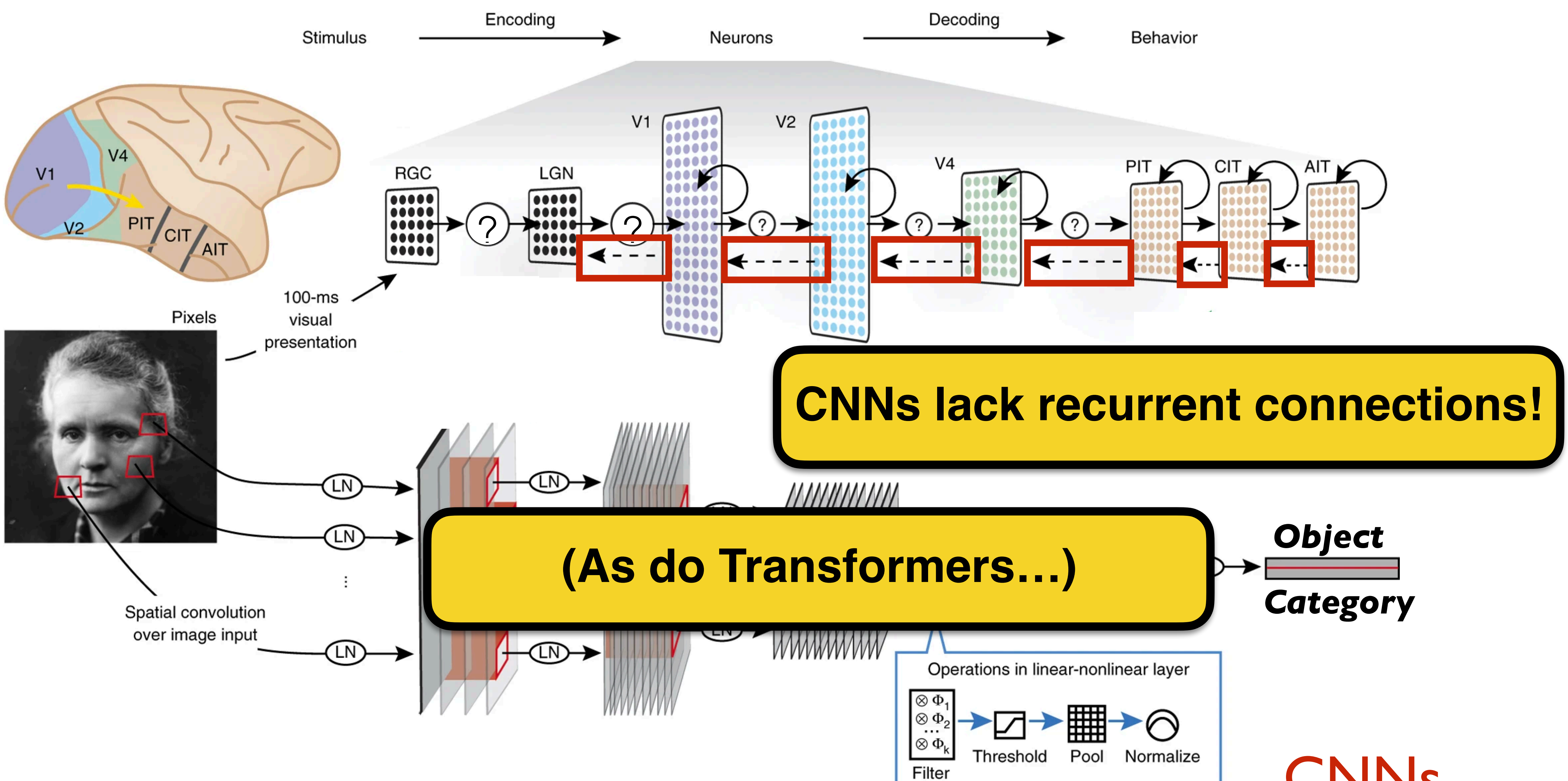
CNNs as Models of Primate Object Recognition



CNNs lack recurrent connections!

(As do Transformers…)

CNNs as Models of Primate Object Recognition

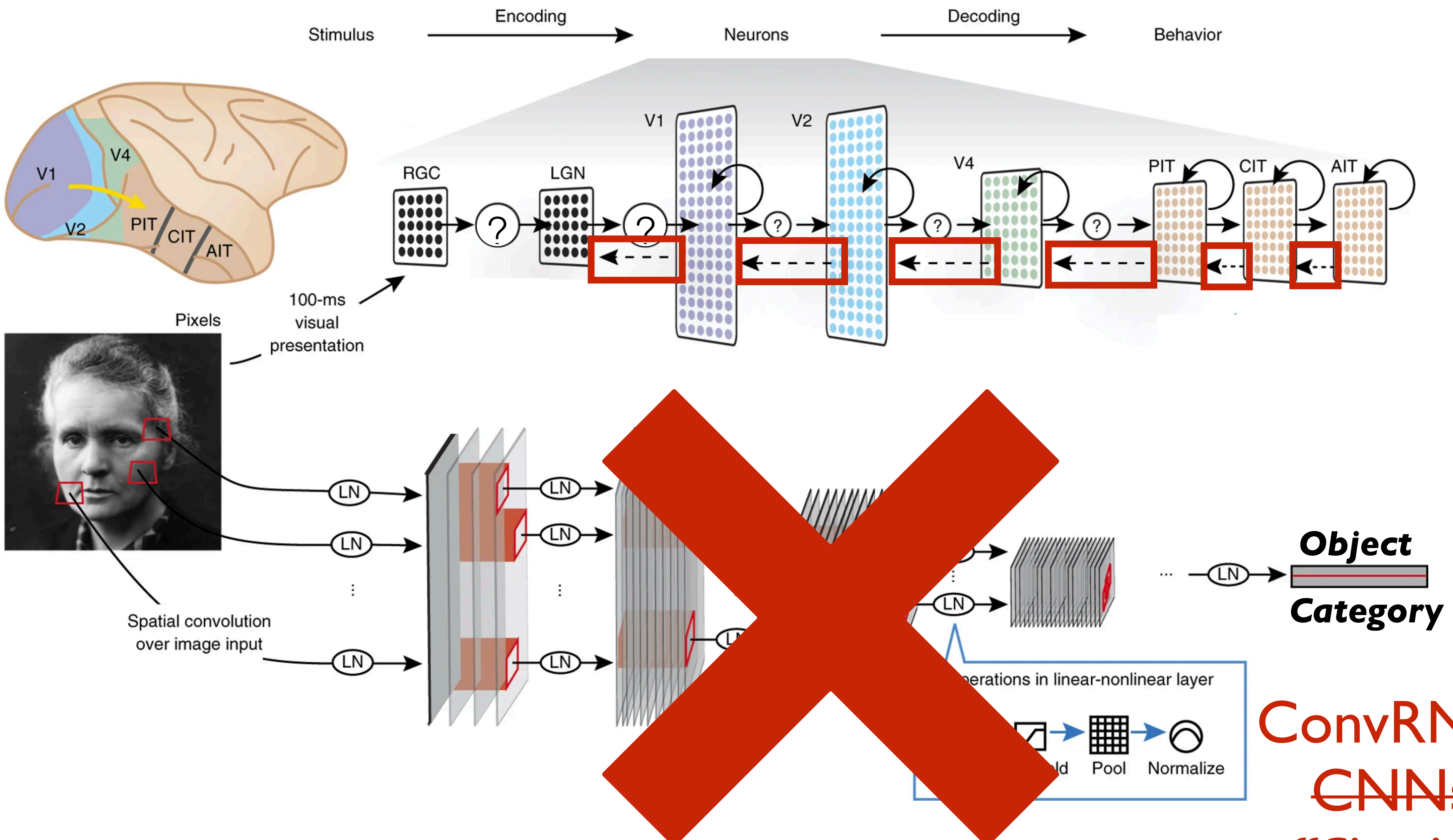


CNNs

"Circuit"

A = *architecture class*

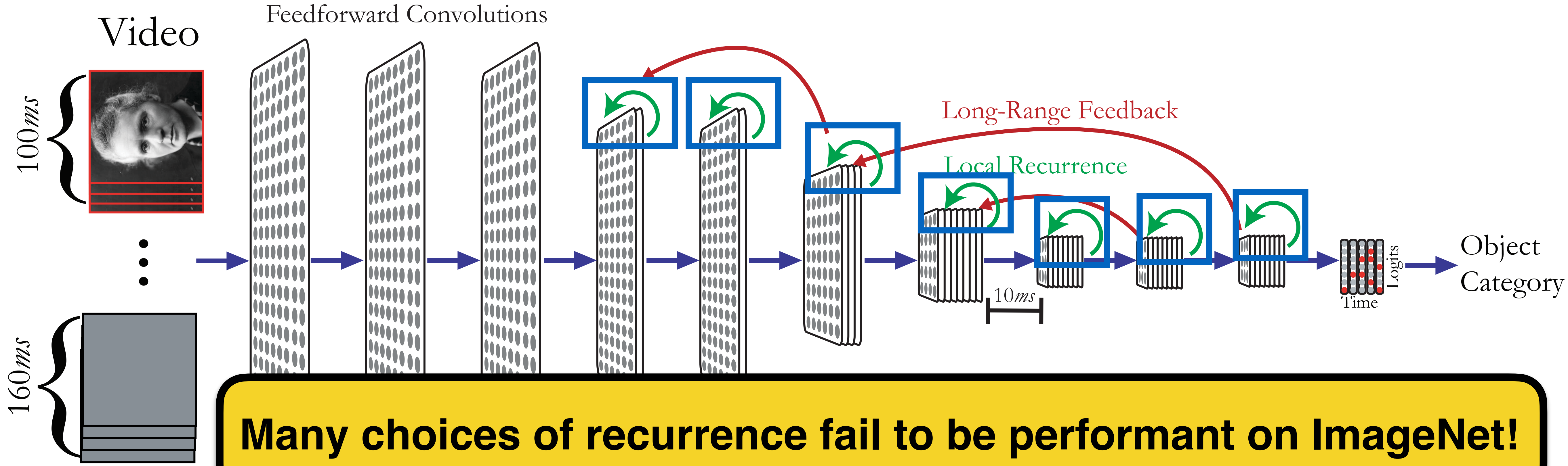
Convolutional Recurrent Networks (ConvRNNs)



ConvRNNs
~~CNNs~~
"Circuit"

A = *architecture class*

Convolutional Recurrent Networks (ConvRNNs)



Many choices of recurrence fail to be performant on ImageNet!

Each time-step (10 *ms*) is treated equally
— including feedforward steps

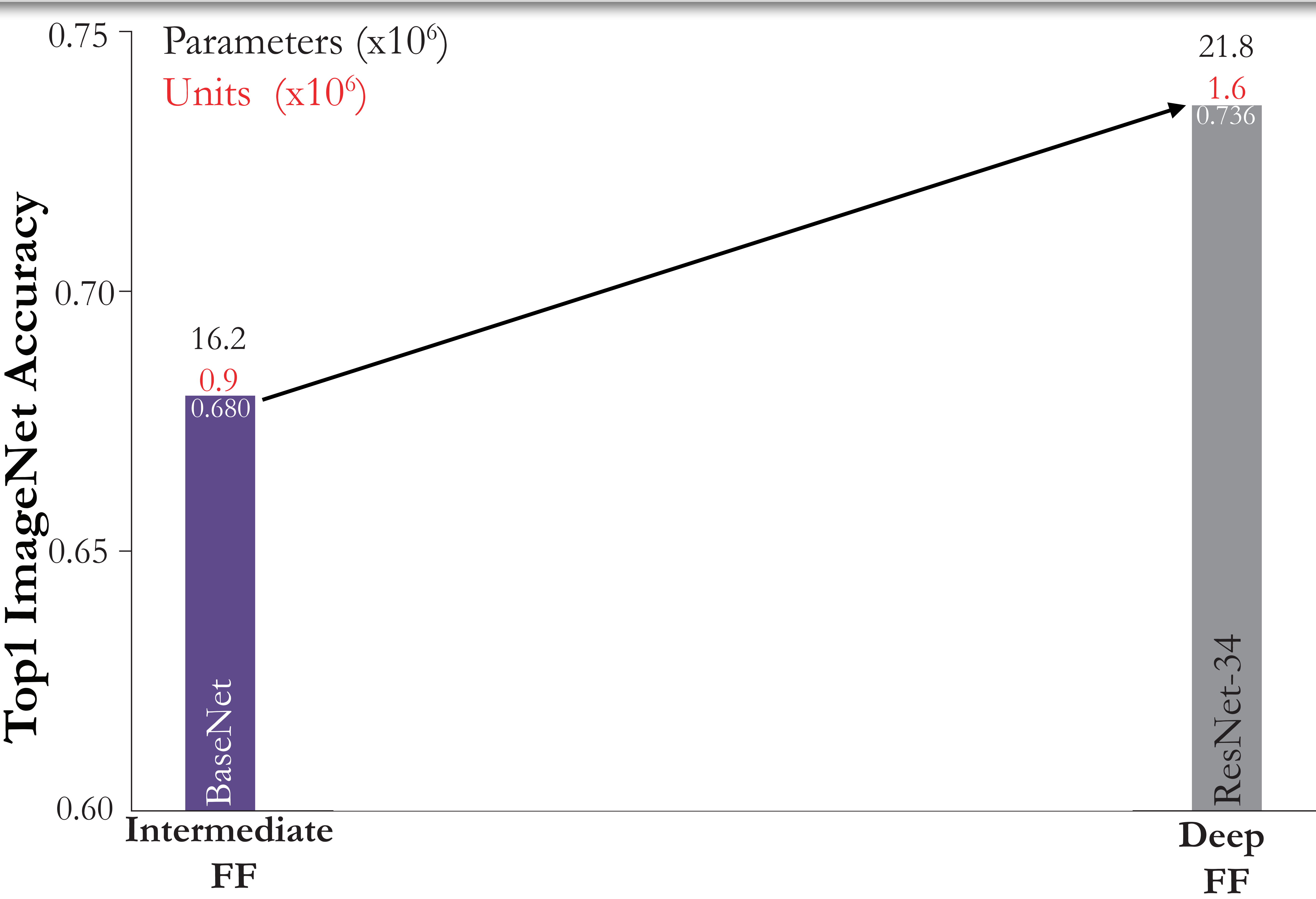
ConvRNNs

~~CNNs~~

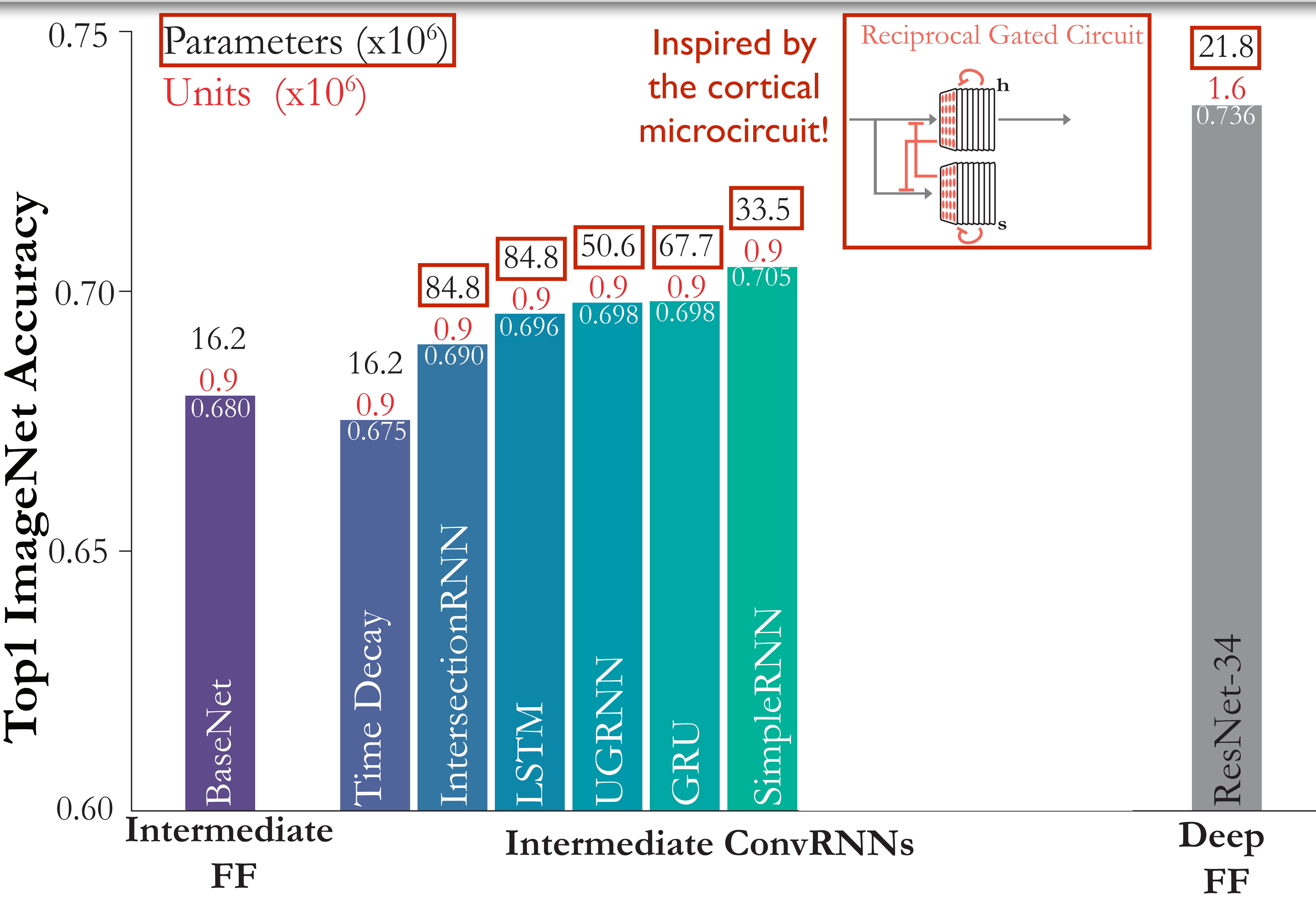
"Circuit"

A = *architecture class*

Implanting Local Recurrence into Feedforward CNNs

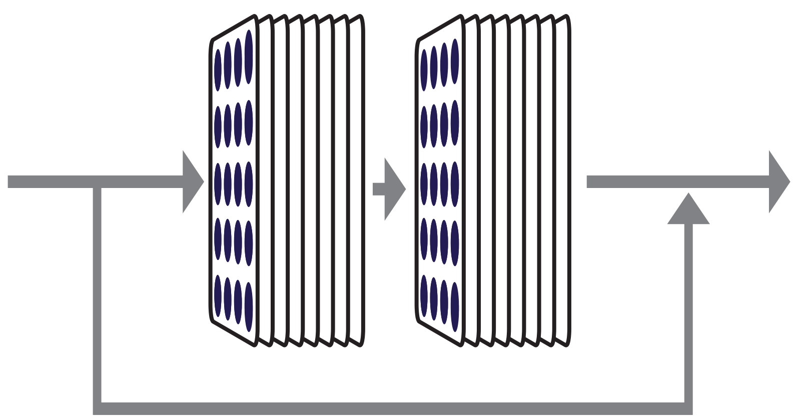


Adding Standard RNNs Helps Incrementally, but Add **Lots** of Parameters!

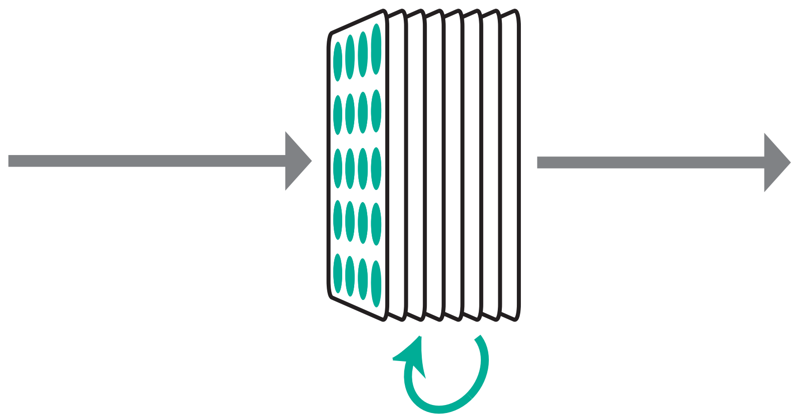


Many Choices of Local Recurrence

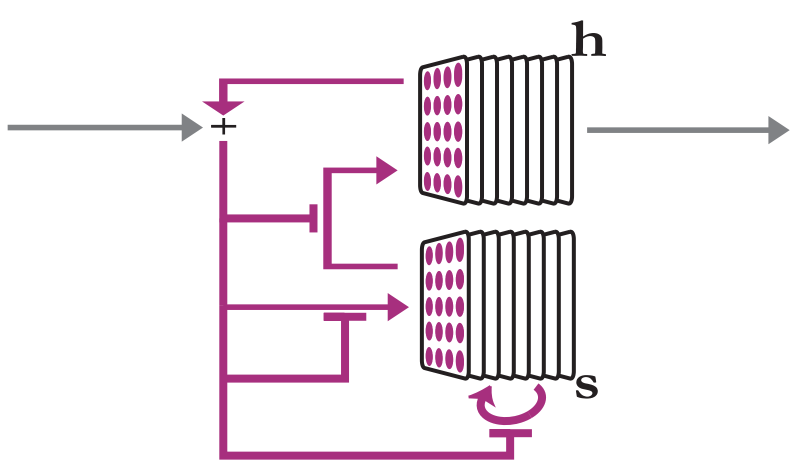
ResNet



SimpleRNN



LSTM

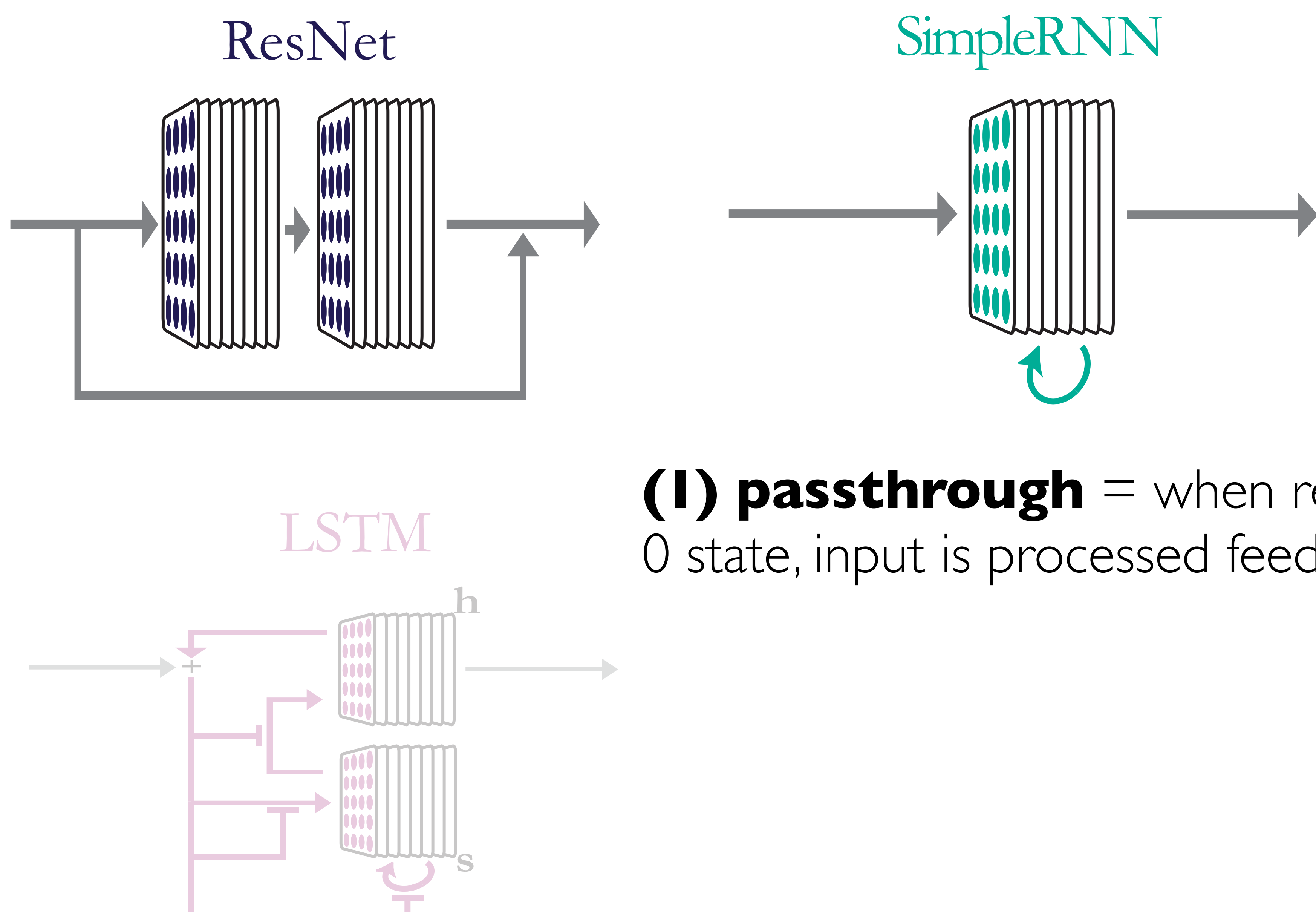


Circuit Diagram

Many Choices of Local Recurrence

State = 0:

Input $x_t \xrightarrow{\text{Passthrough Mechanism}}$ Output $f(W*x_t+b)$



(1) passthrough = when recurrent cell is in 0 state, input is processed feedforward as usual

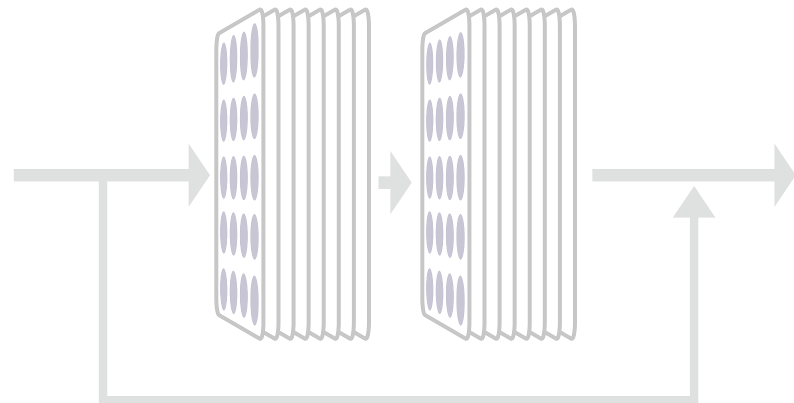
Circuit Diagram

Many Choices of Local Recurrence

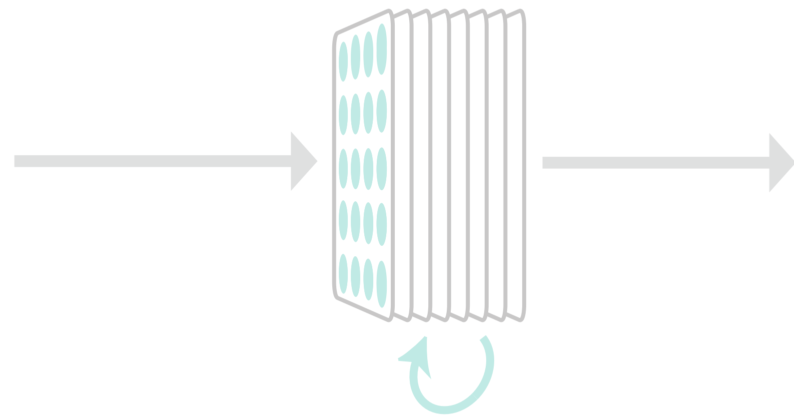
Passthrough Mechanism — *State = 0:* Input x_t → Output $f(W*x_t+b)$

Gating Mechanism — Input x_t → Output $f(W*x_t+b) \circ g_t$, g_t

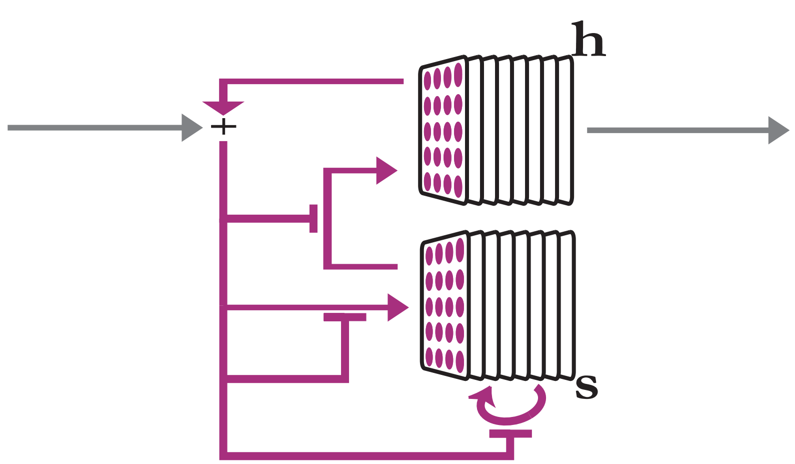
ResNet



SimpleRNN



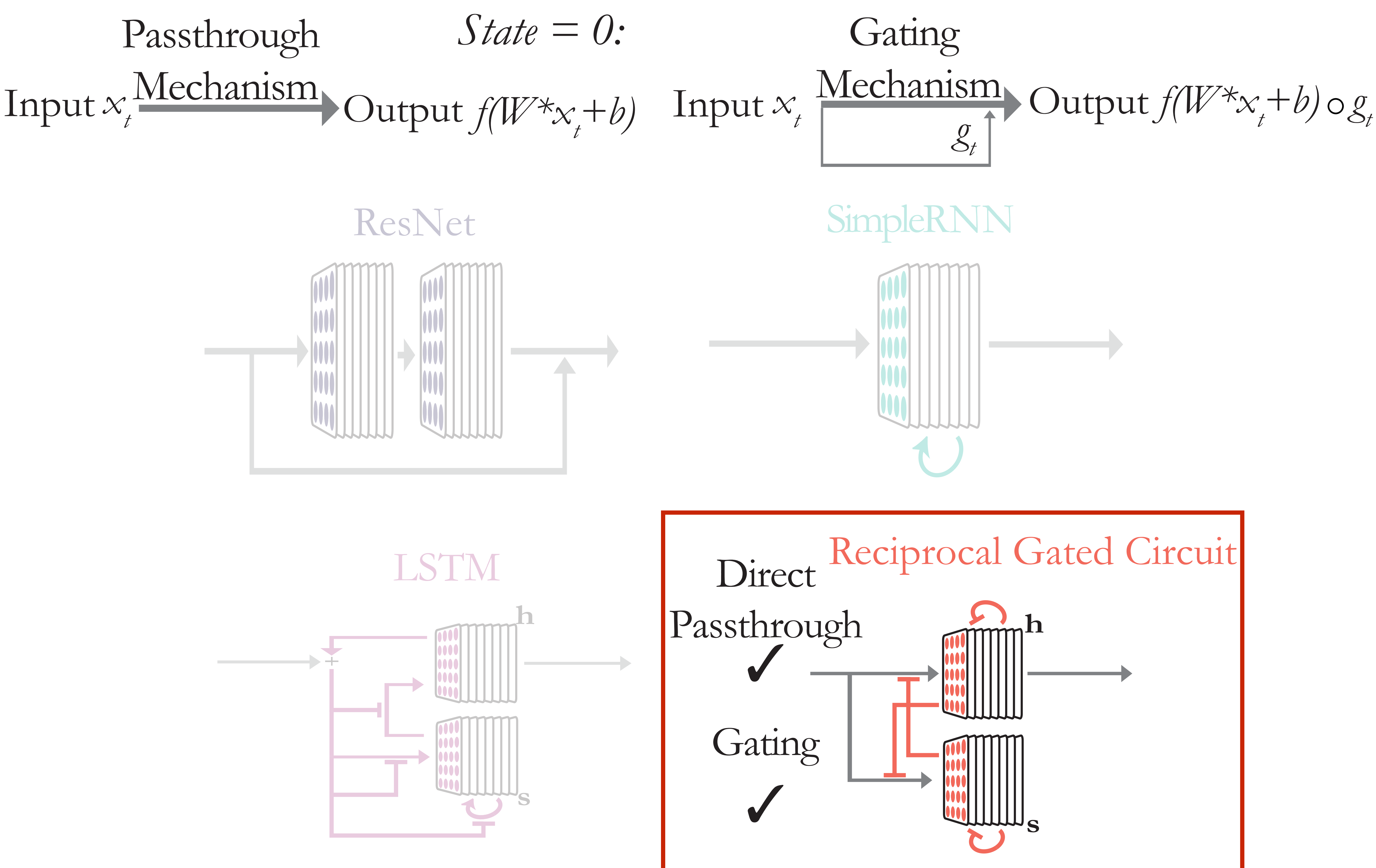
LSTM



Circuit Diagram

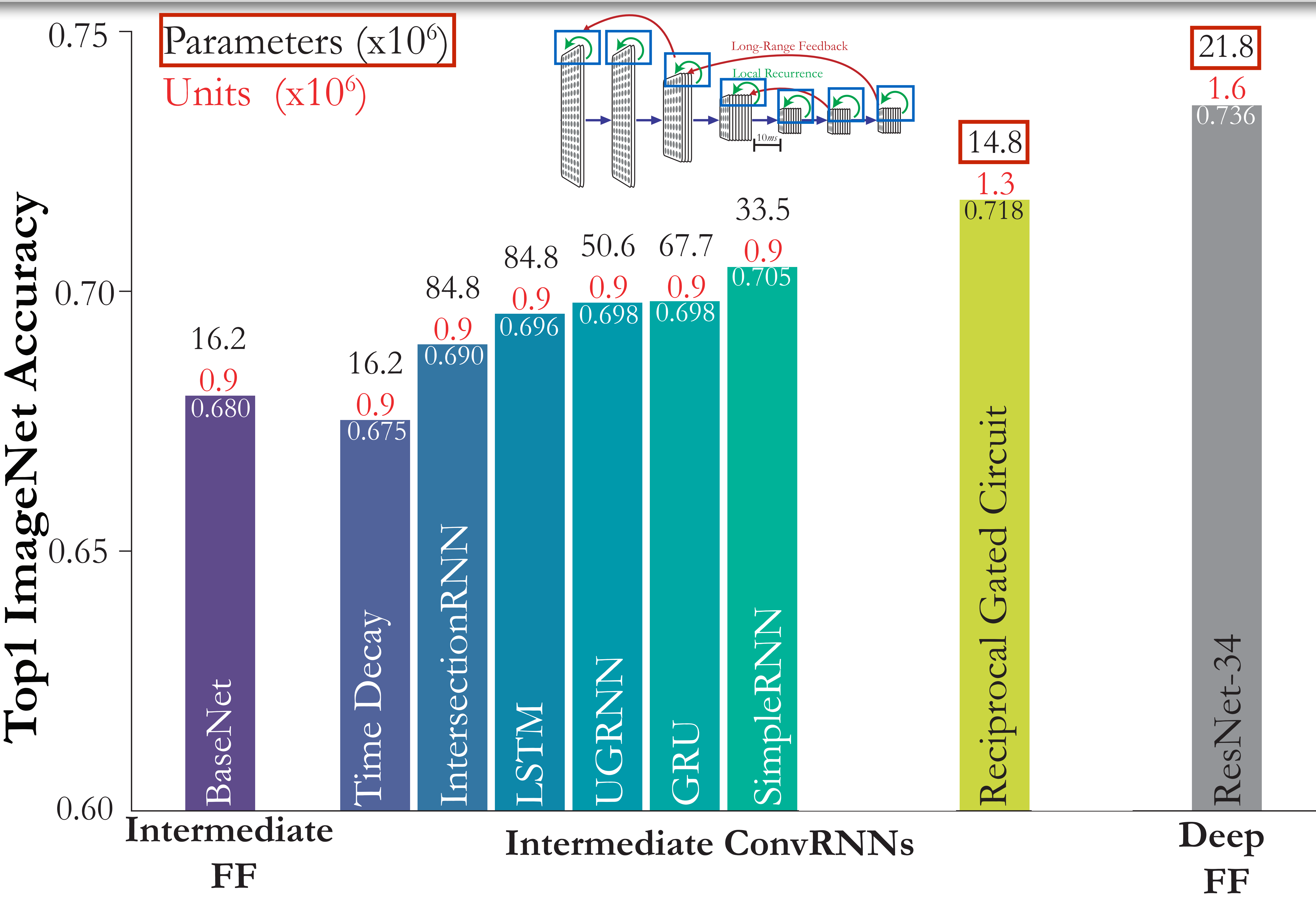
(2) gating = multiplication by input-dependent tensor w/ values in [0, 1]

Many Choices of Local Recurrence

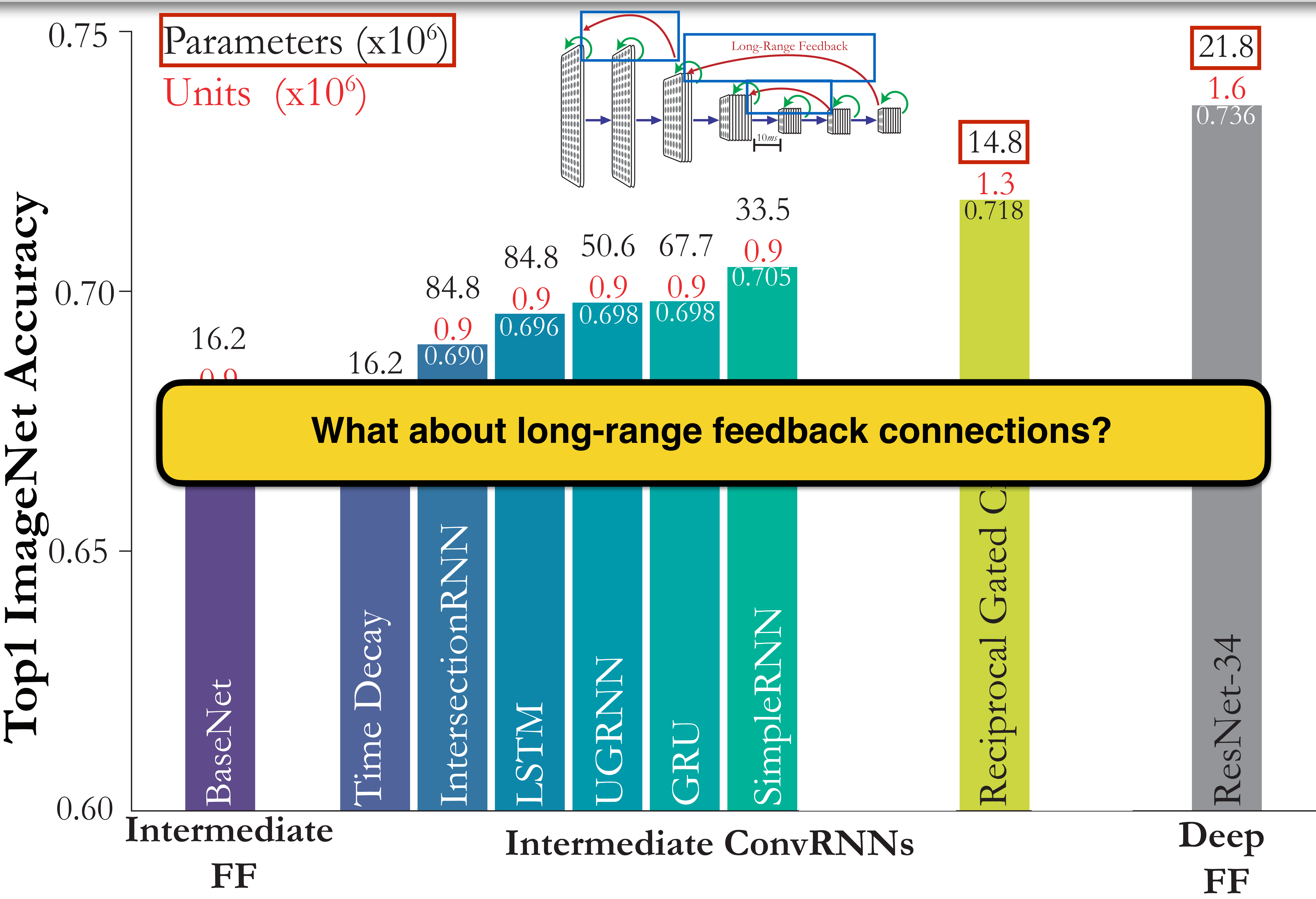


Circuit Diagram

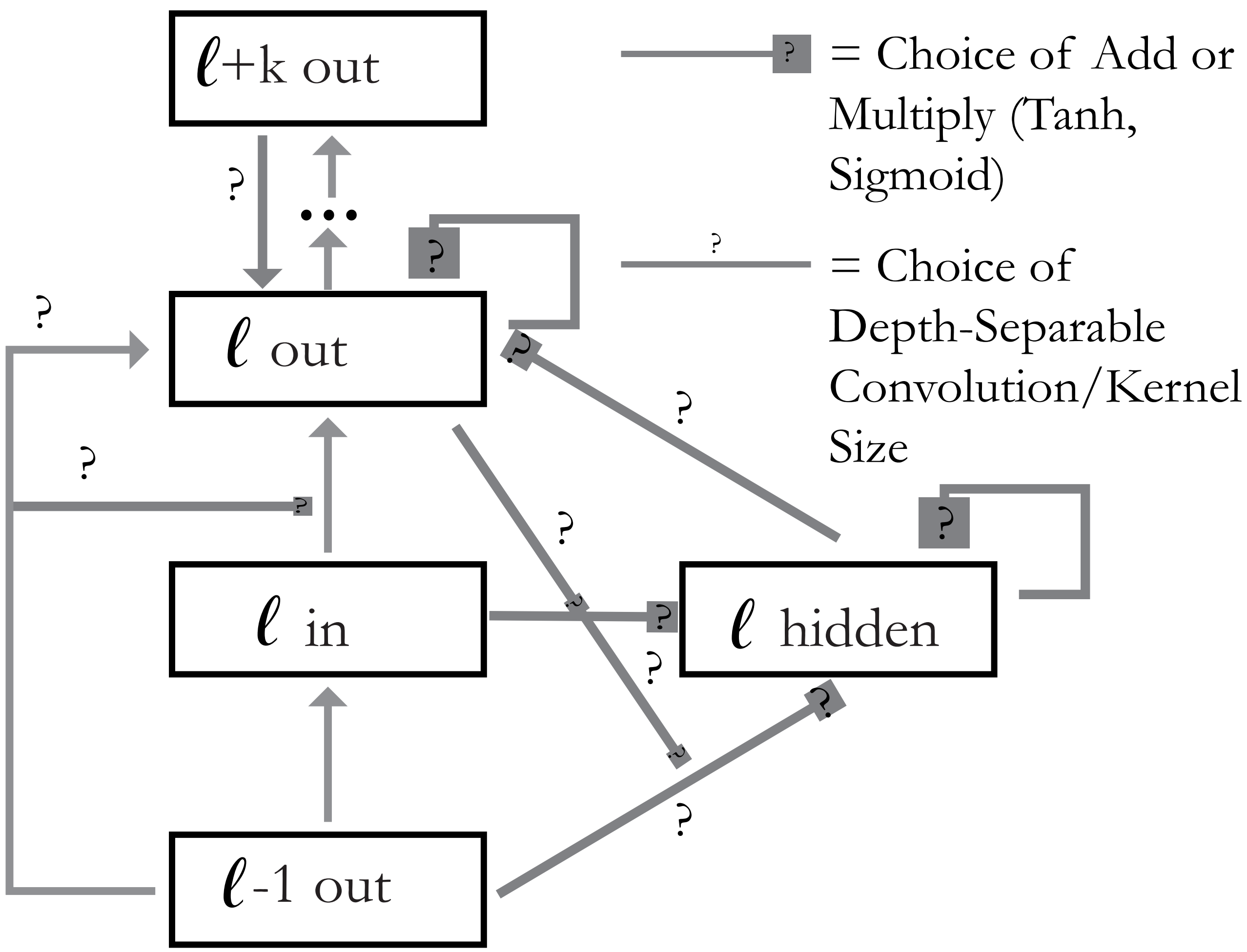
Novel Recurrent Cells Yield Improved ImageNet Performance

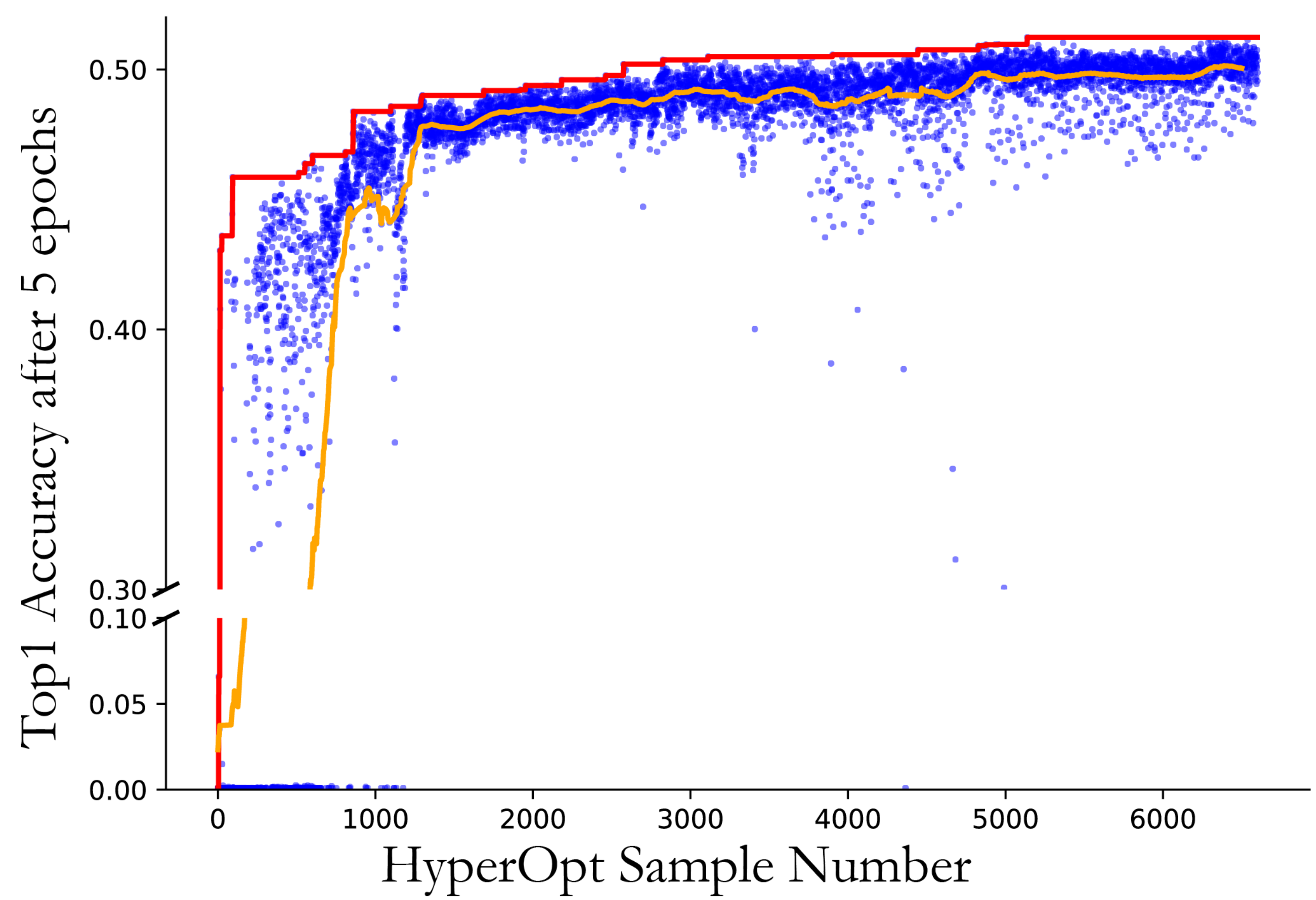


Novel Recurrent Cells Yield Improved ImageNet Performance

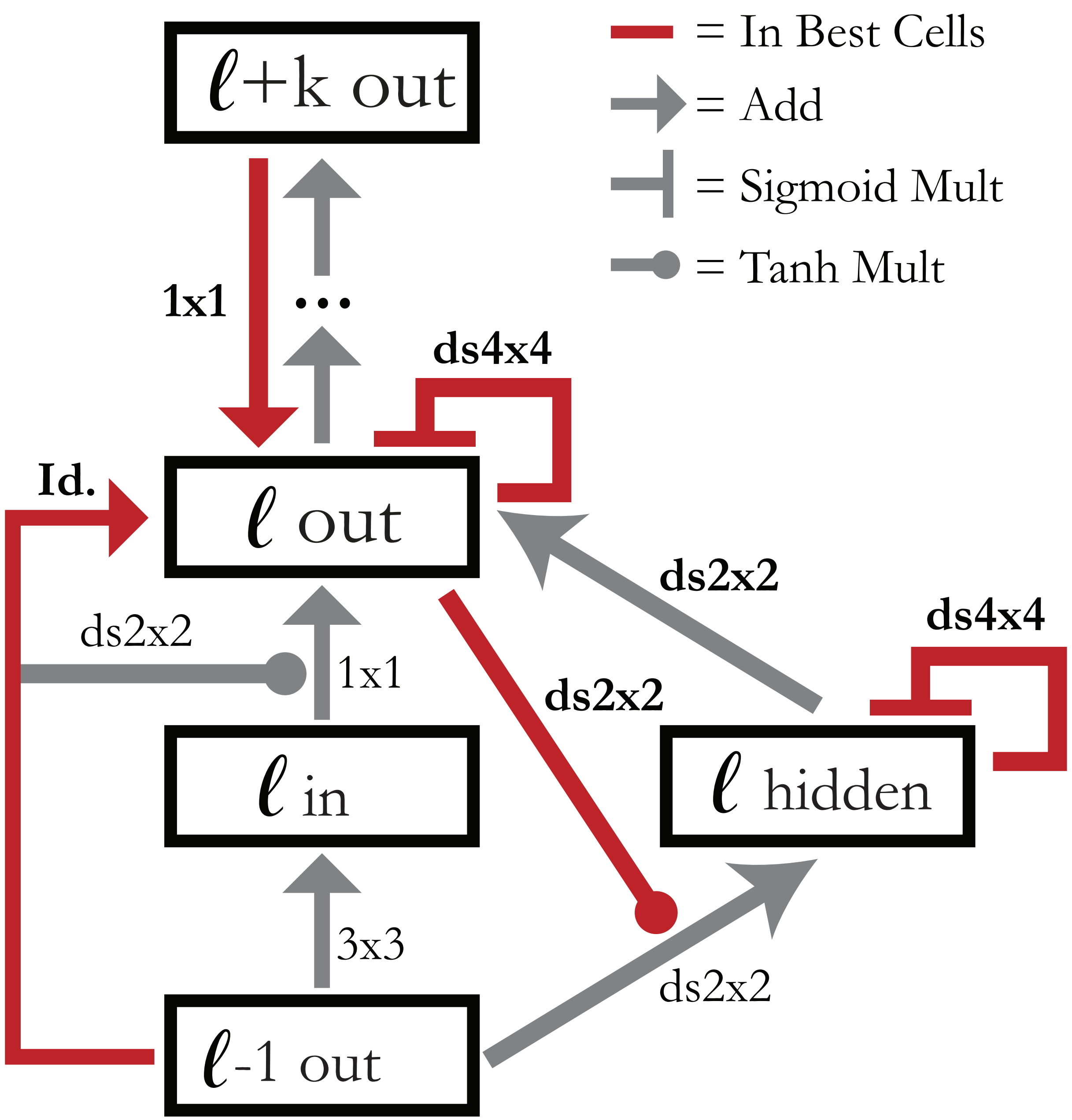


Large-Scale Search Over Long-Range Feedback Connections



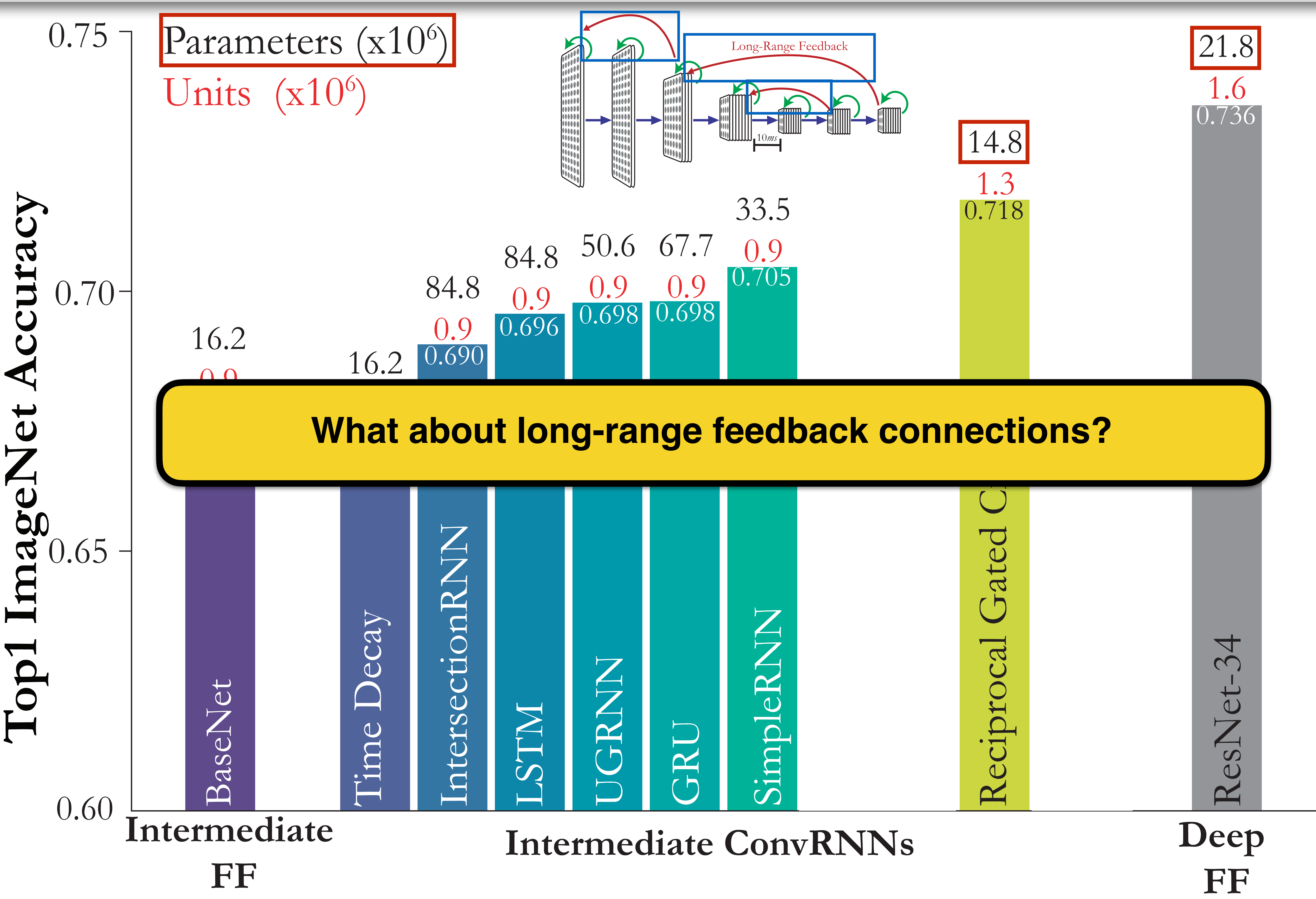


Emergent Global Connectivity Patterns



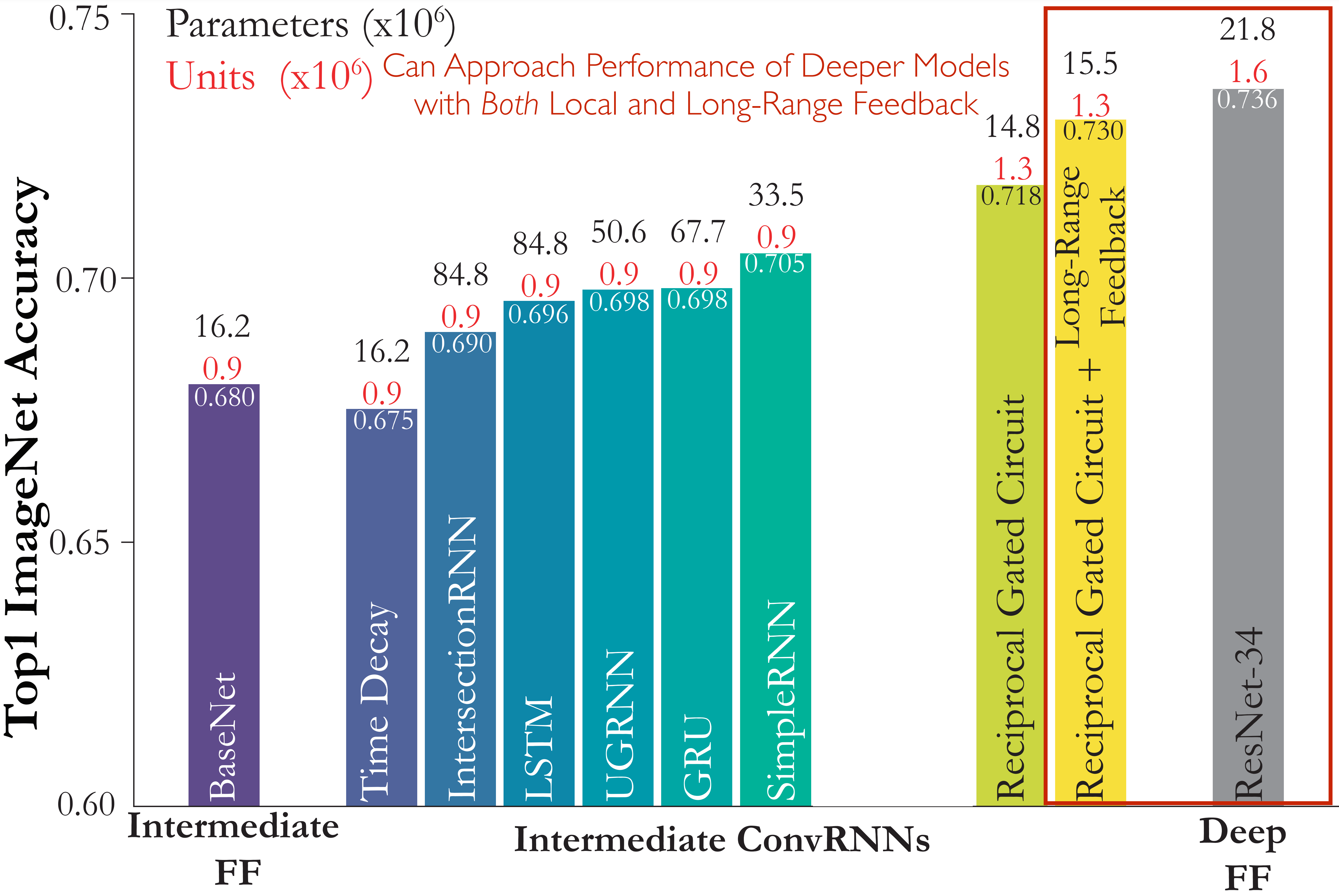
Conservation on parameter count as a byproduct of evolutionary optimization

Novel Recurrent Cells Yield Improved ImageNet Performance

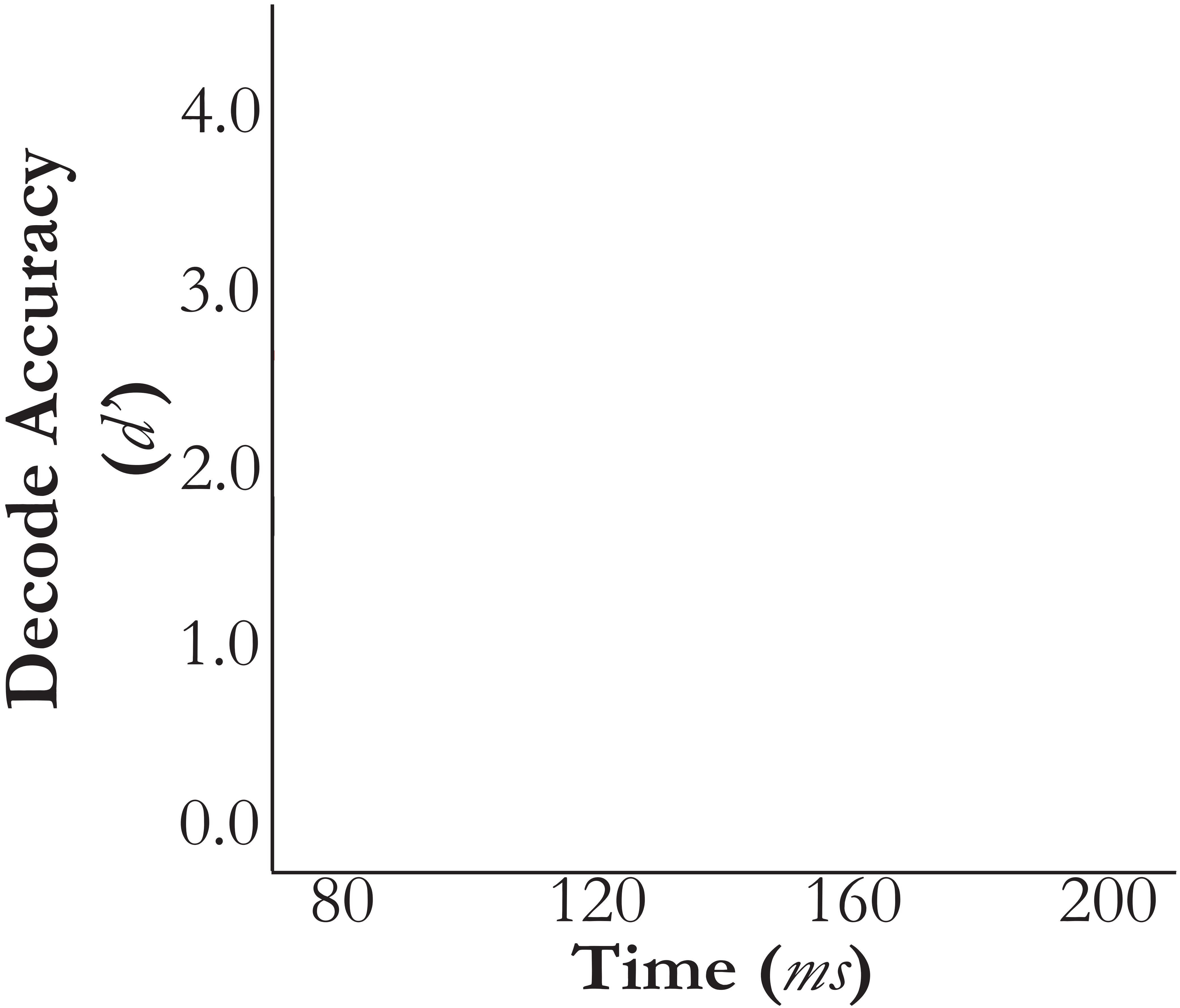


What about long-range feedback connections?

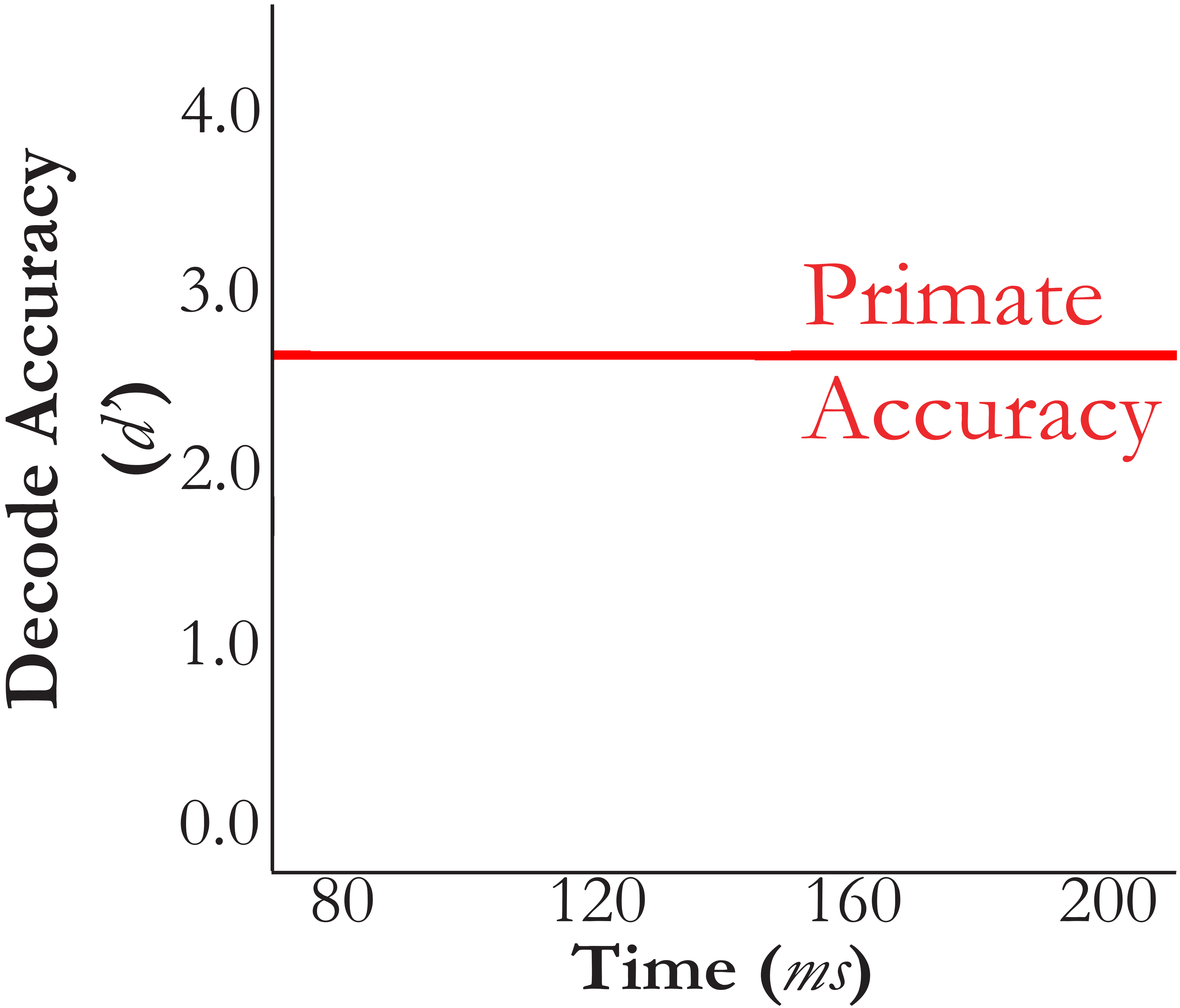
Long-Range Feedback Connections Matter



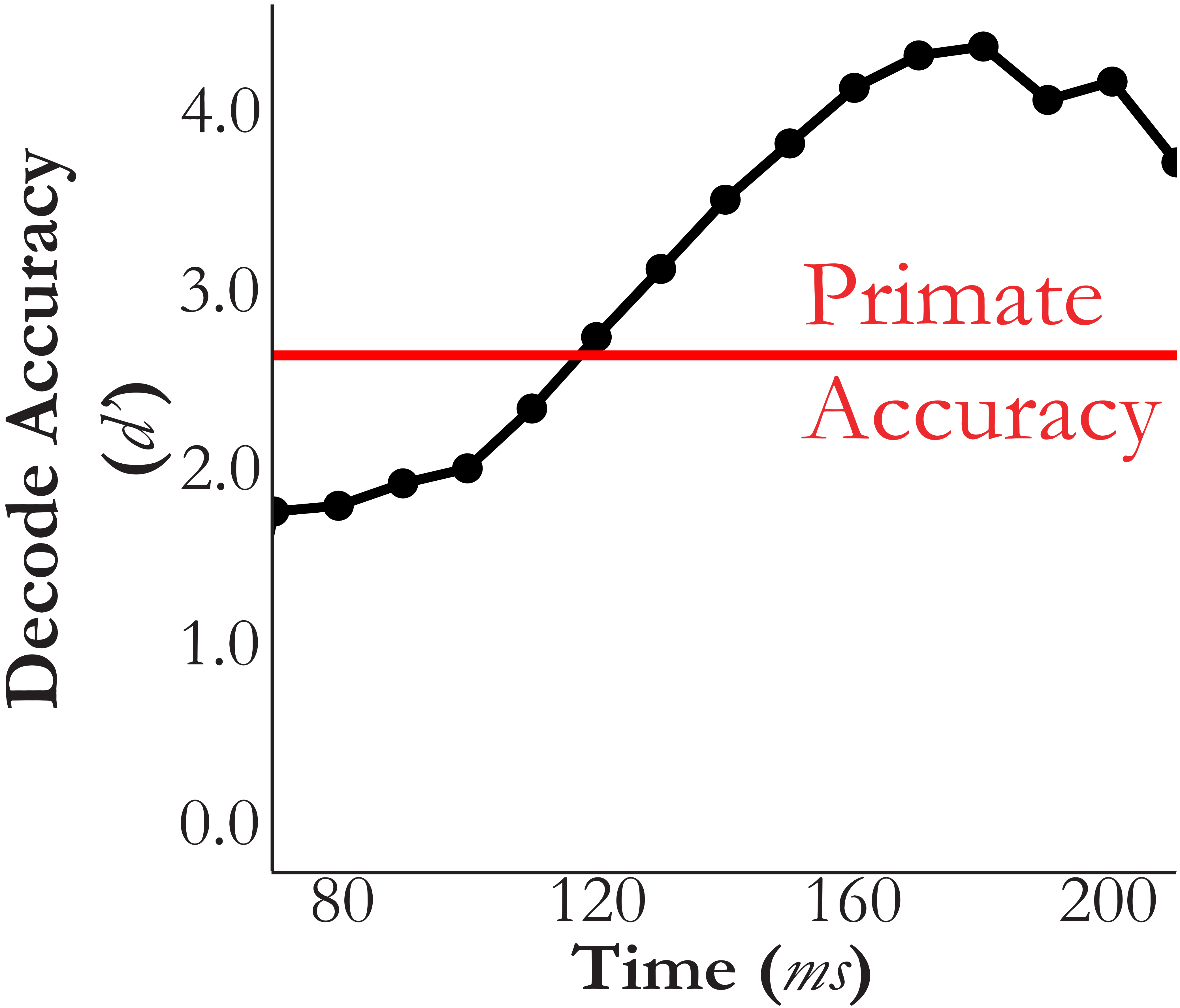
Comparing to Primate Object Solution Times (OSTs)



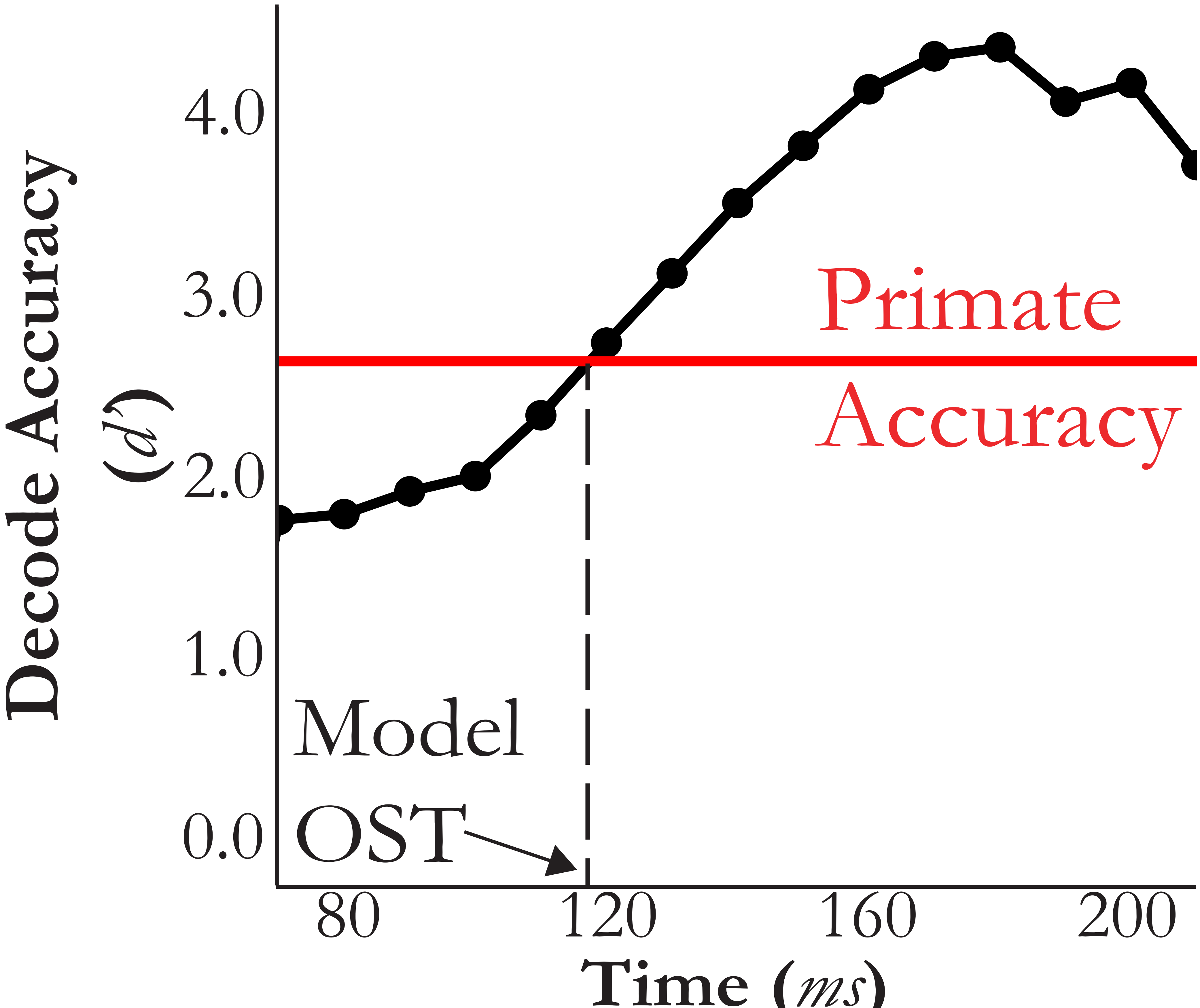
Comparing to Primate Object Solution Times (OSTs)



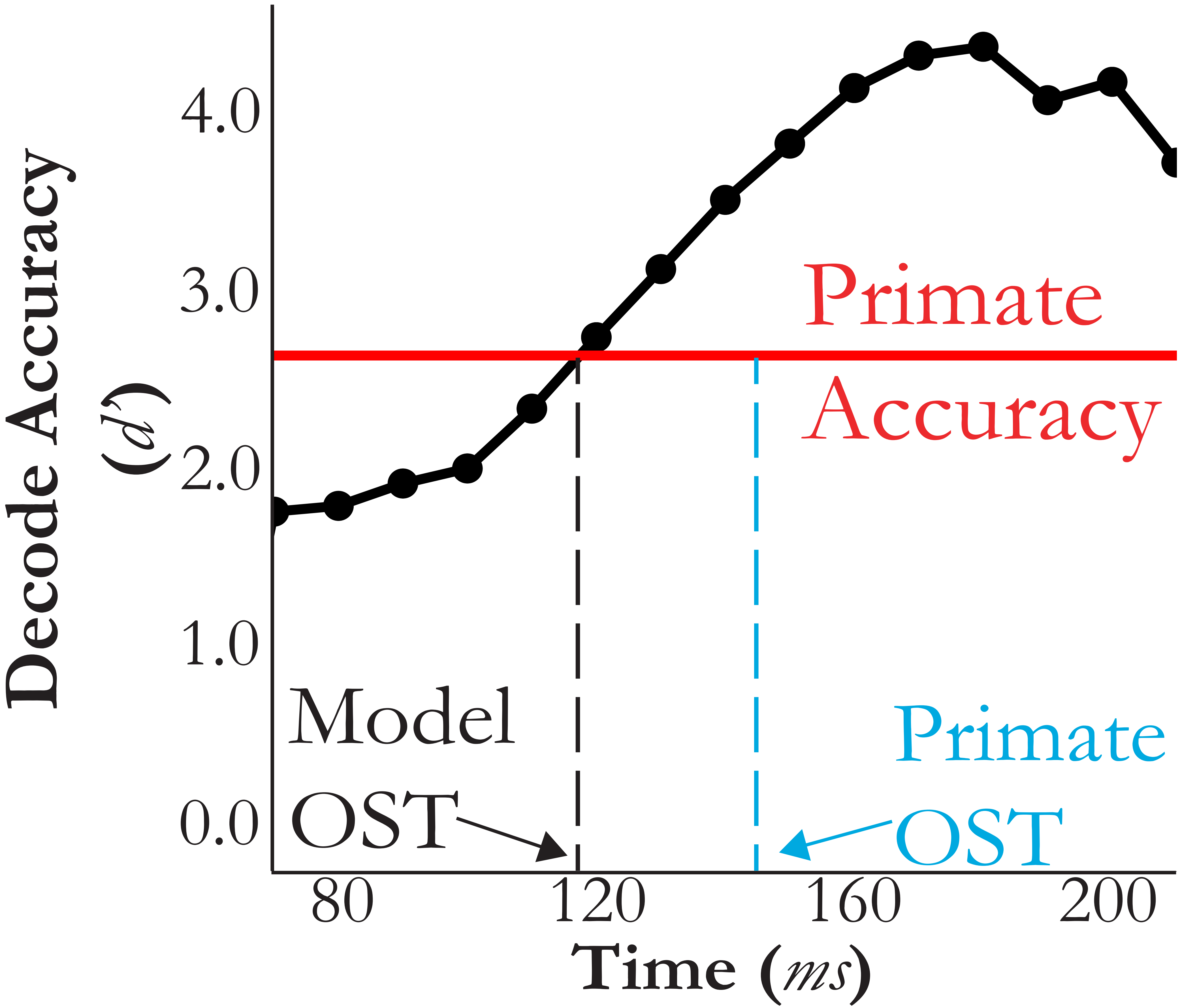
Comparing to Primate Object Solution Times (OSTs)



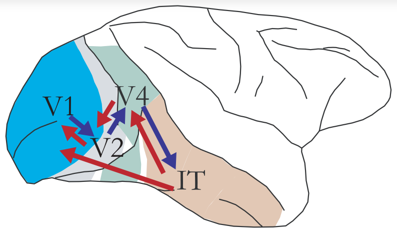
Comparing to Primate Object Solution Times (OSTs)

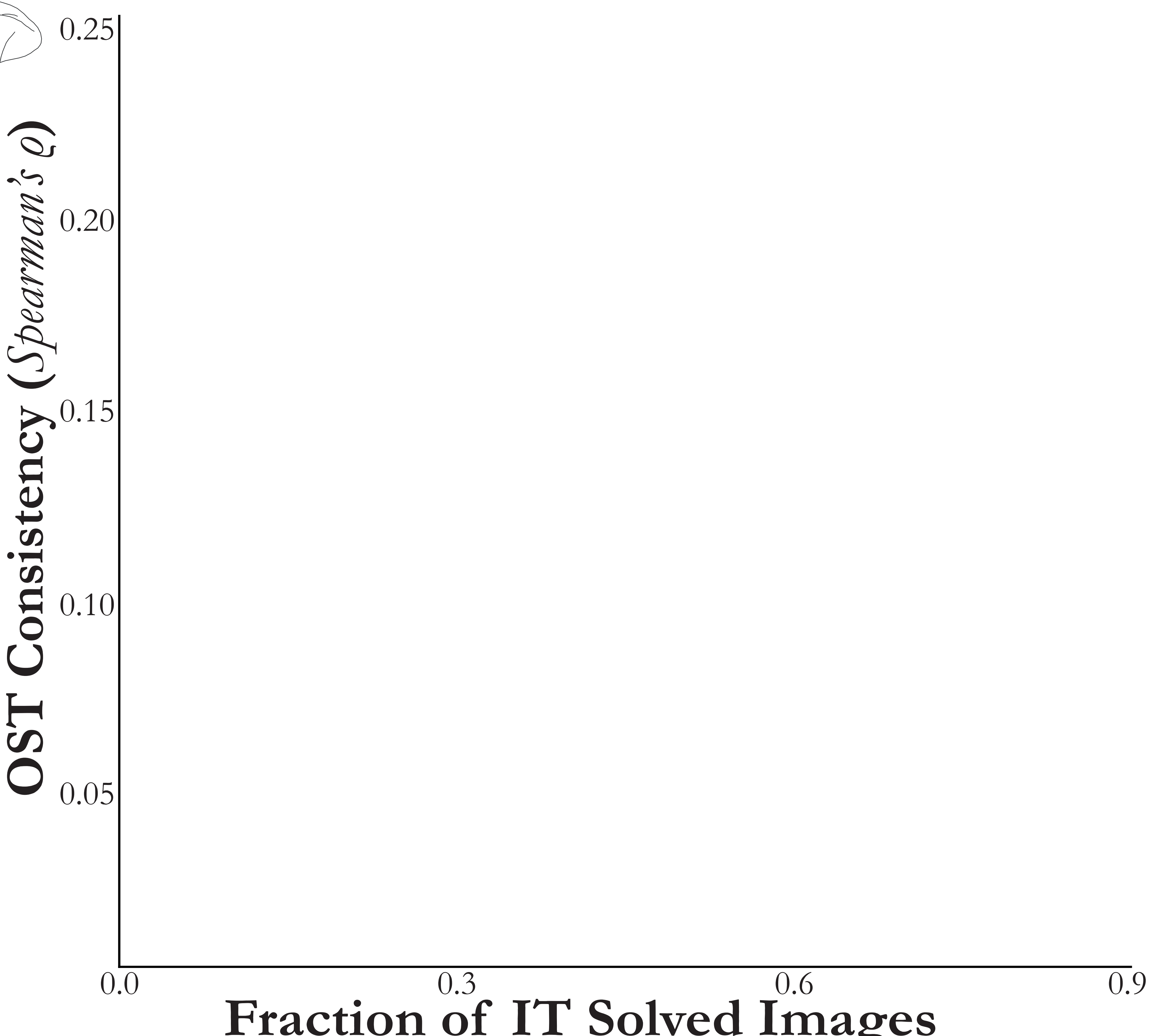


Comparing to Primate Object Solution Times (OSTs)

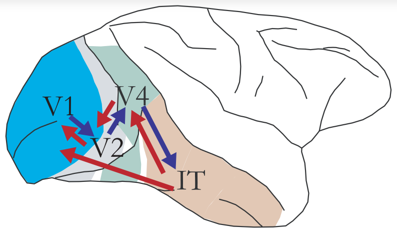


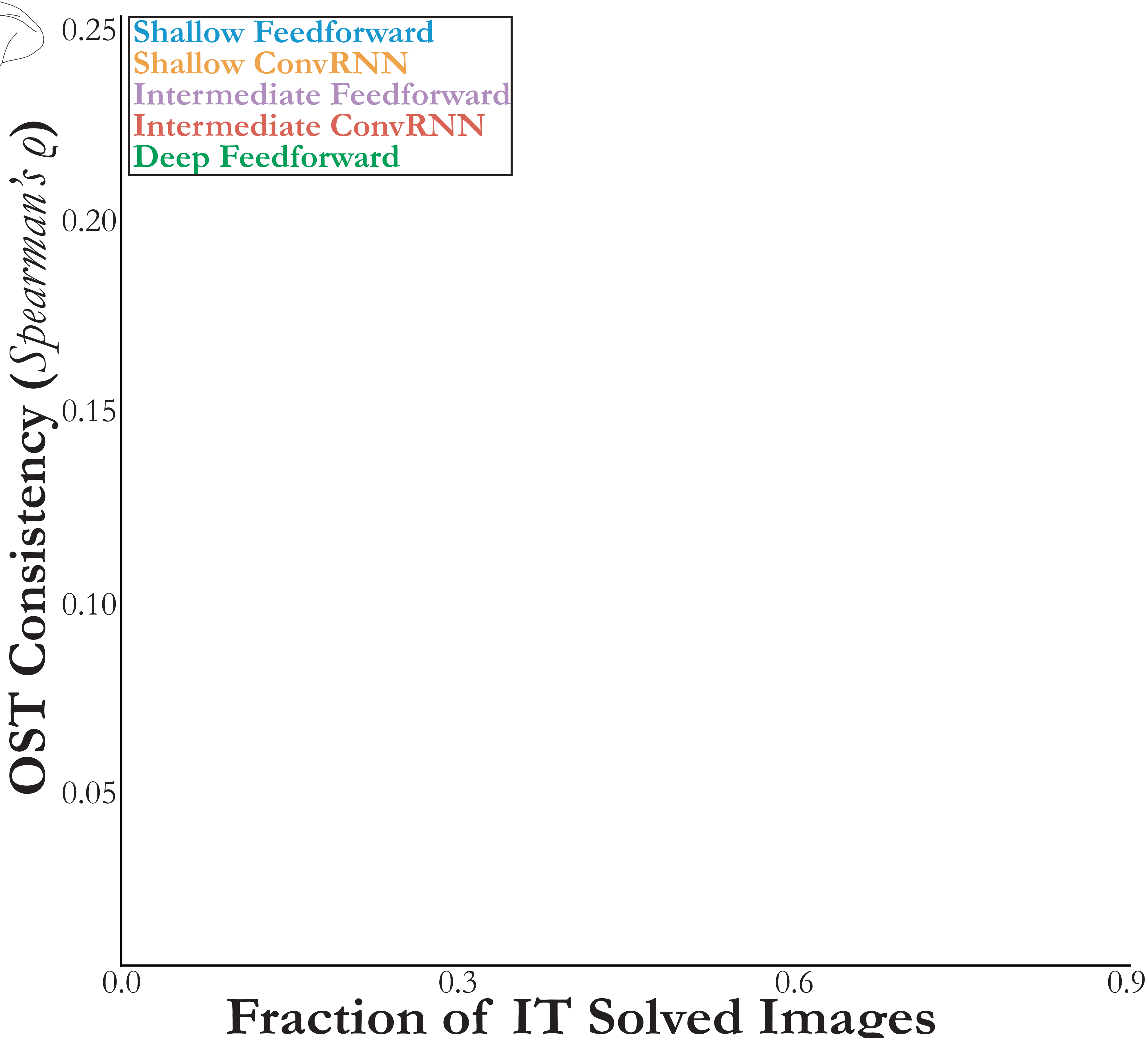
Comparing to Primate Object Solution Times (OSTs)



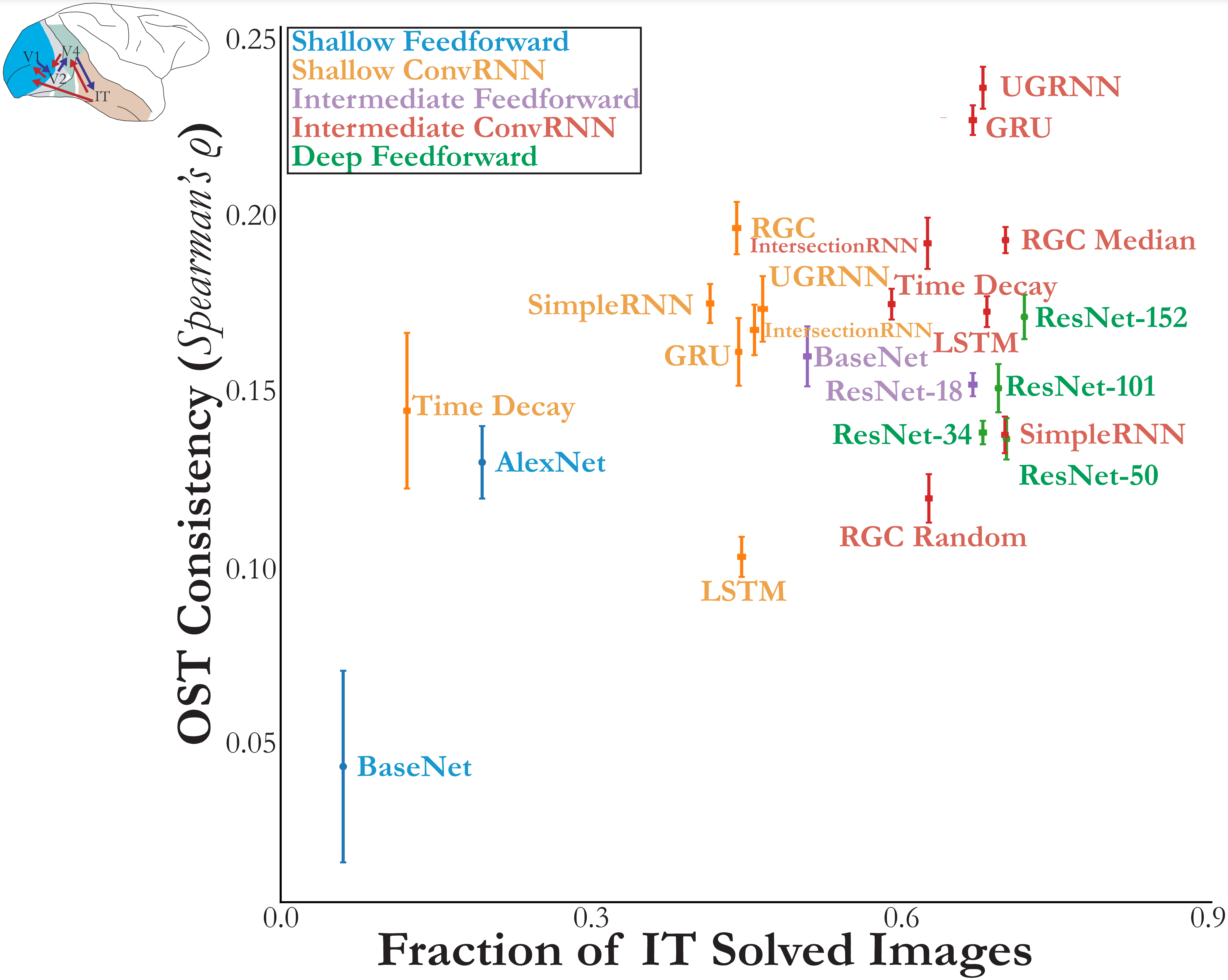


Comparing to Primate Object Solution Times (OSTs)

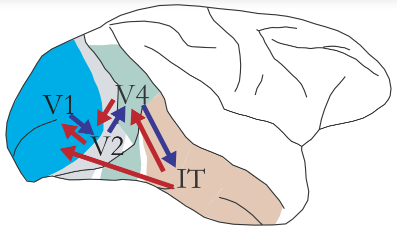


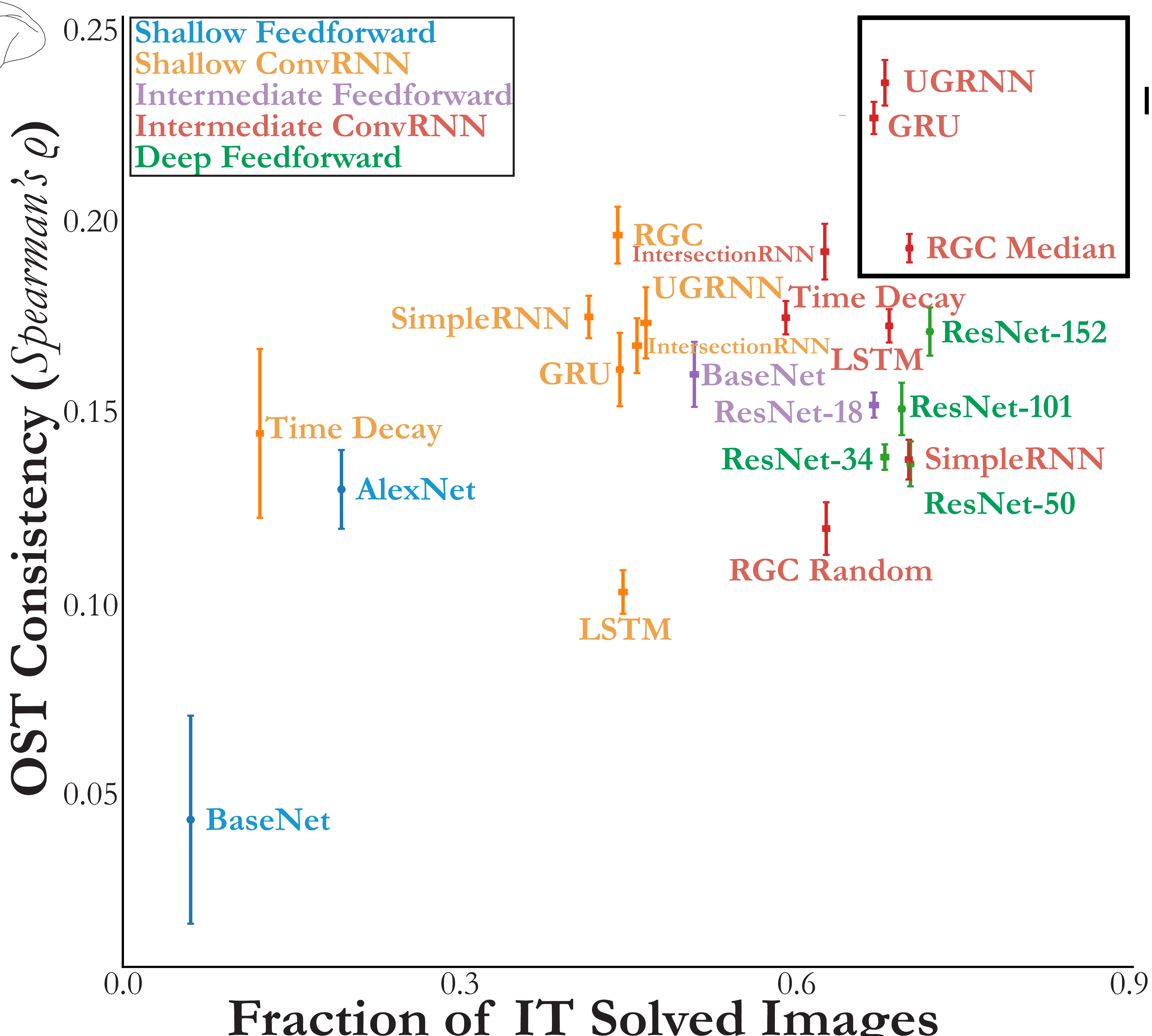


Comparing to Primate Object Solution Times (OSTs)



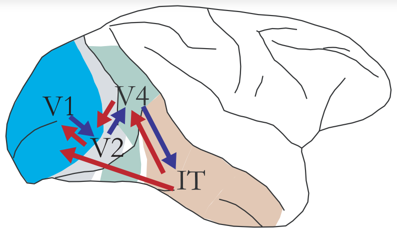
Intermediate Depth ConvRNNs best match OSTs

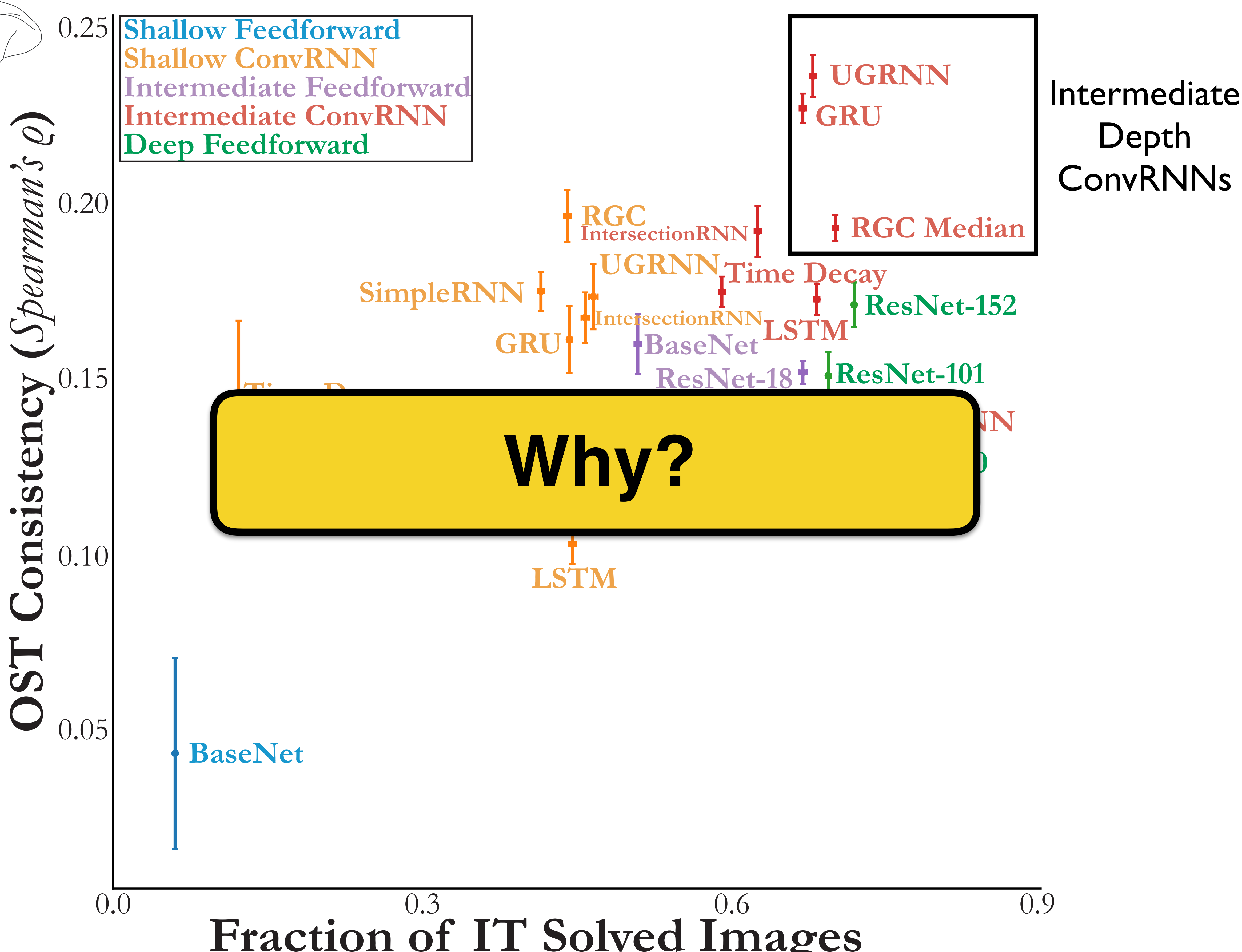




Intermediate Depth ConvRNNs

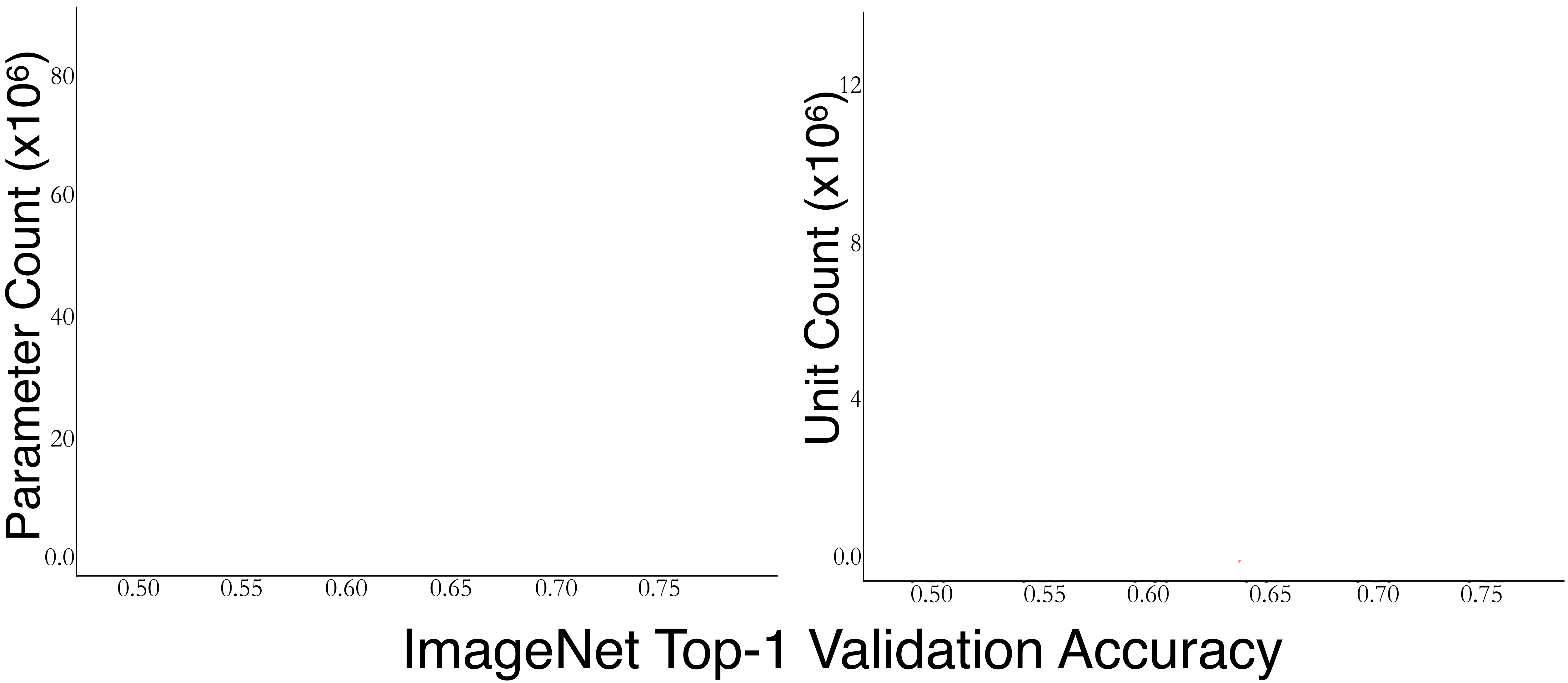
Intermediate Depth ConvRNNs best match OSTs



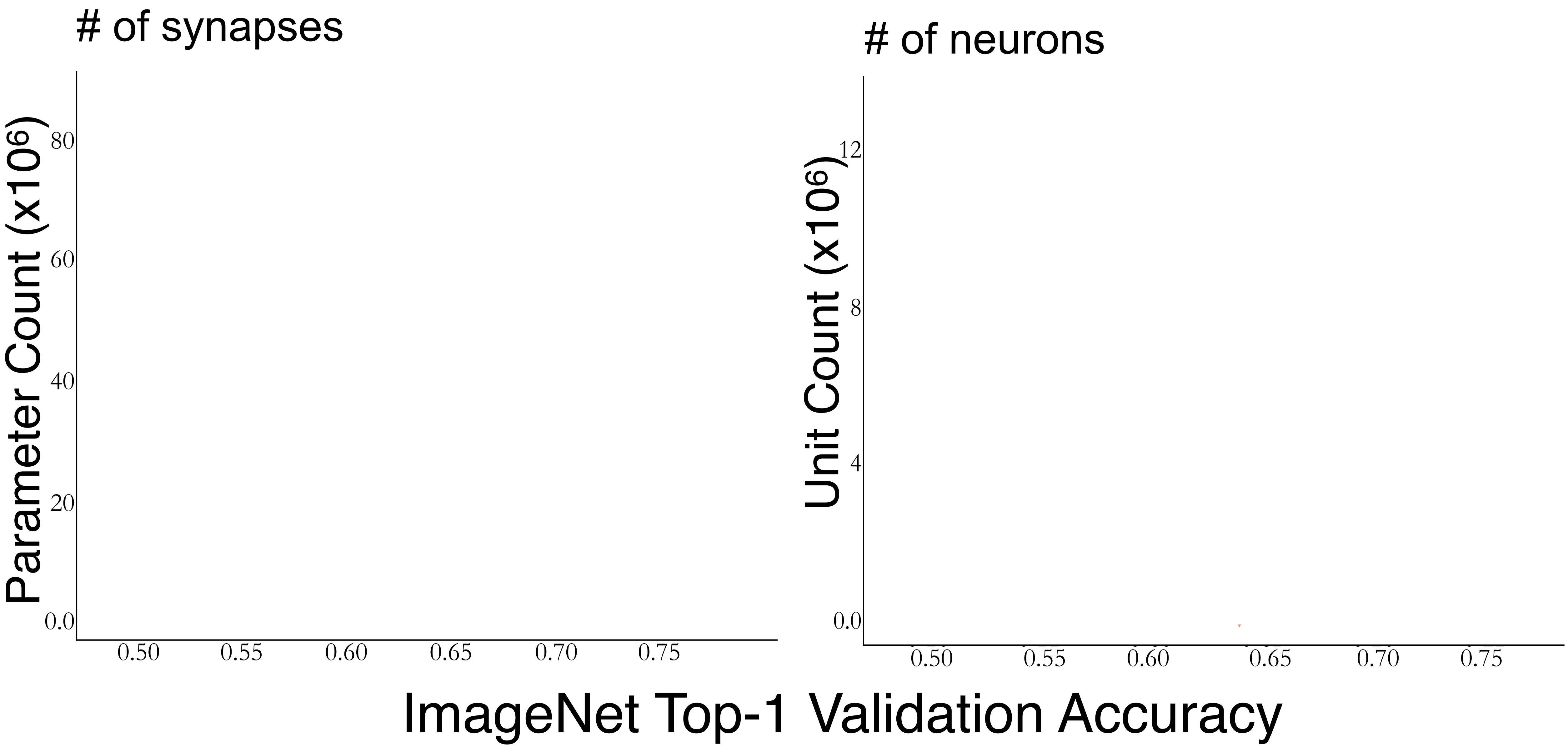


Why?

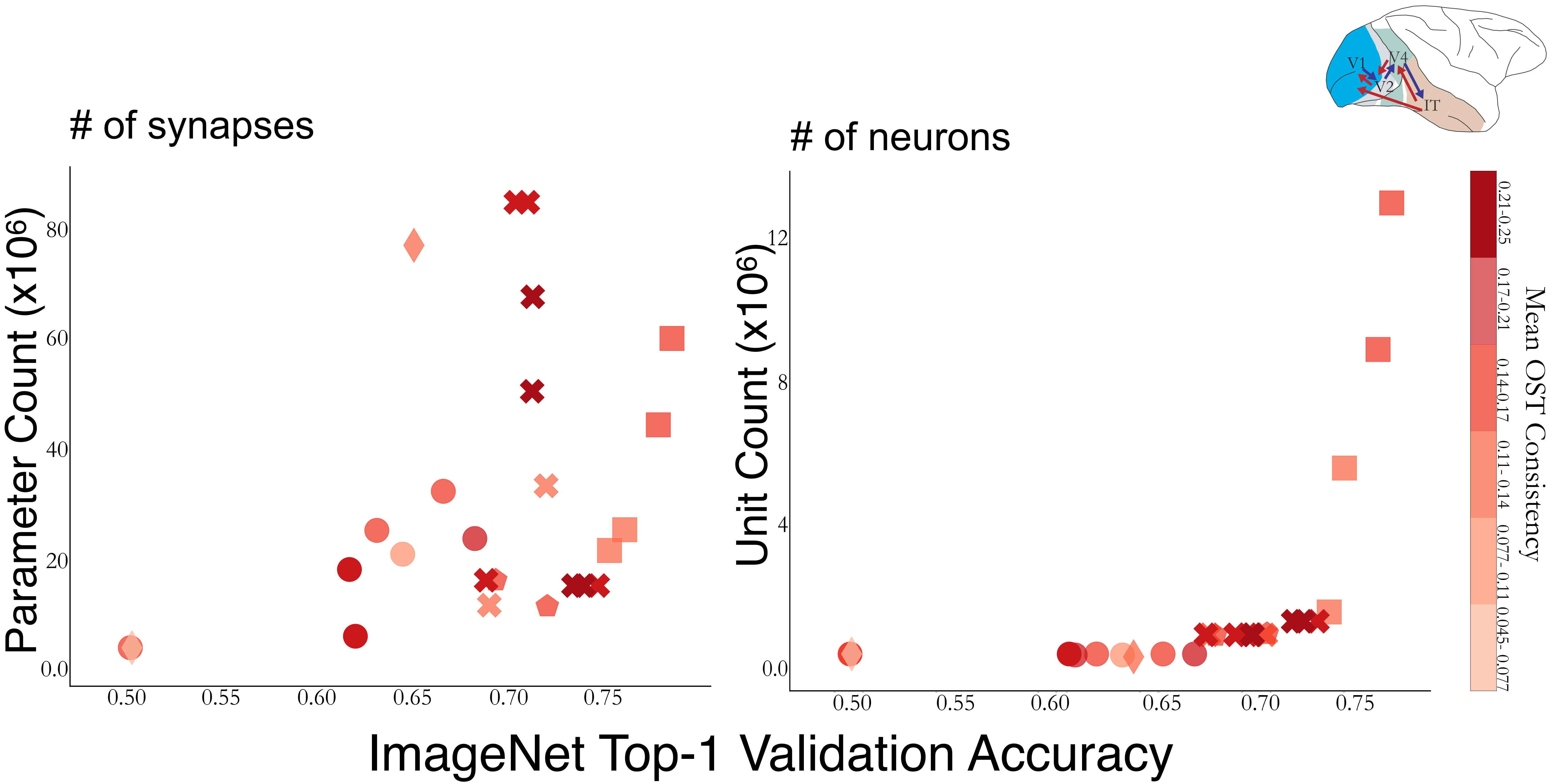
Why Are ConvRNNs the Most Brain-Like?



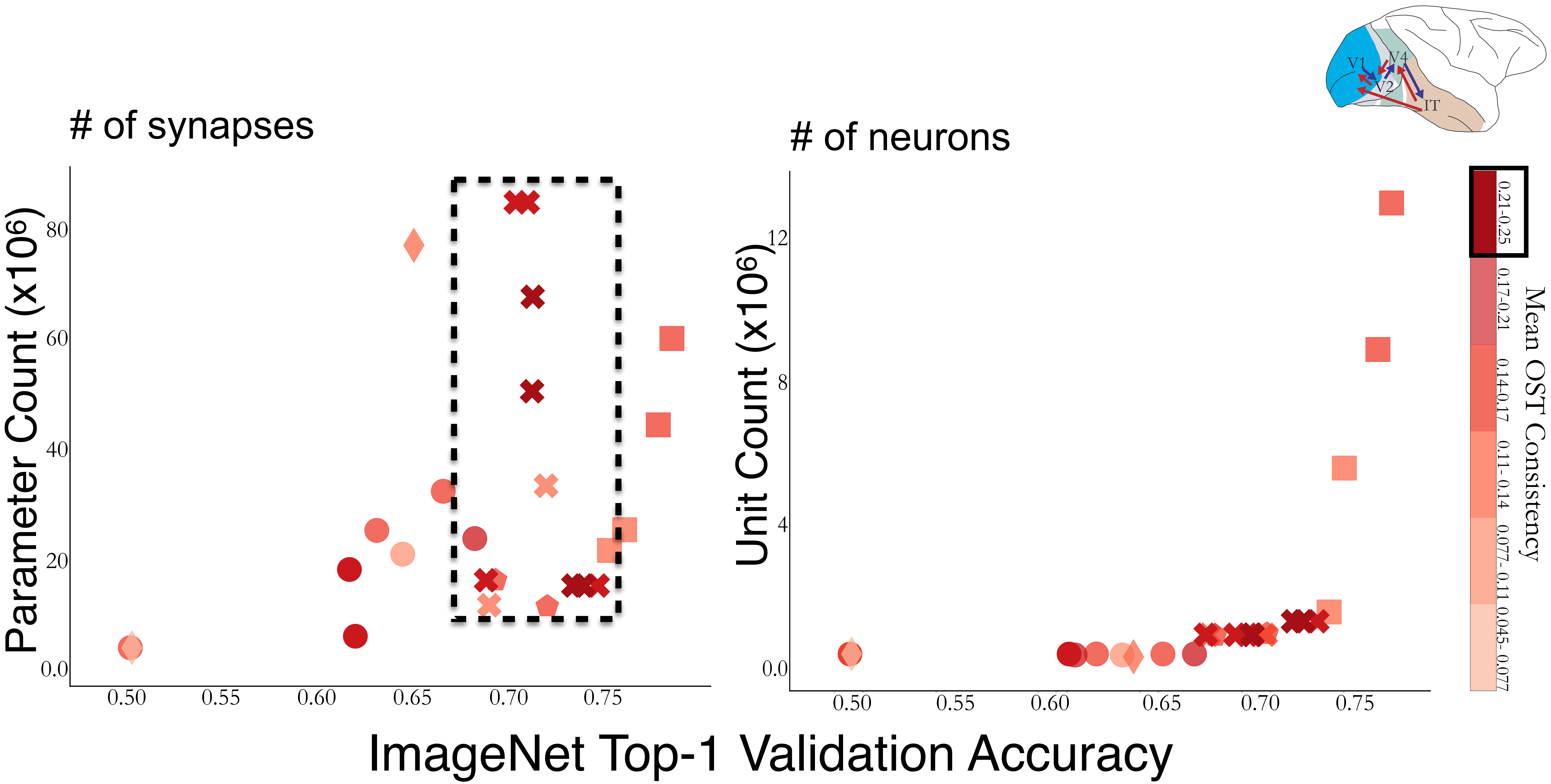
Why Are ConvRNNs the Most Brain-Like?



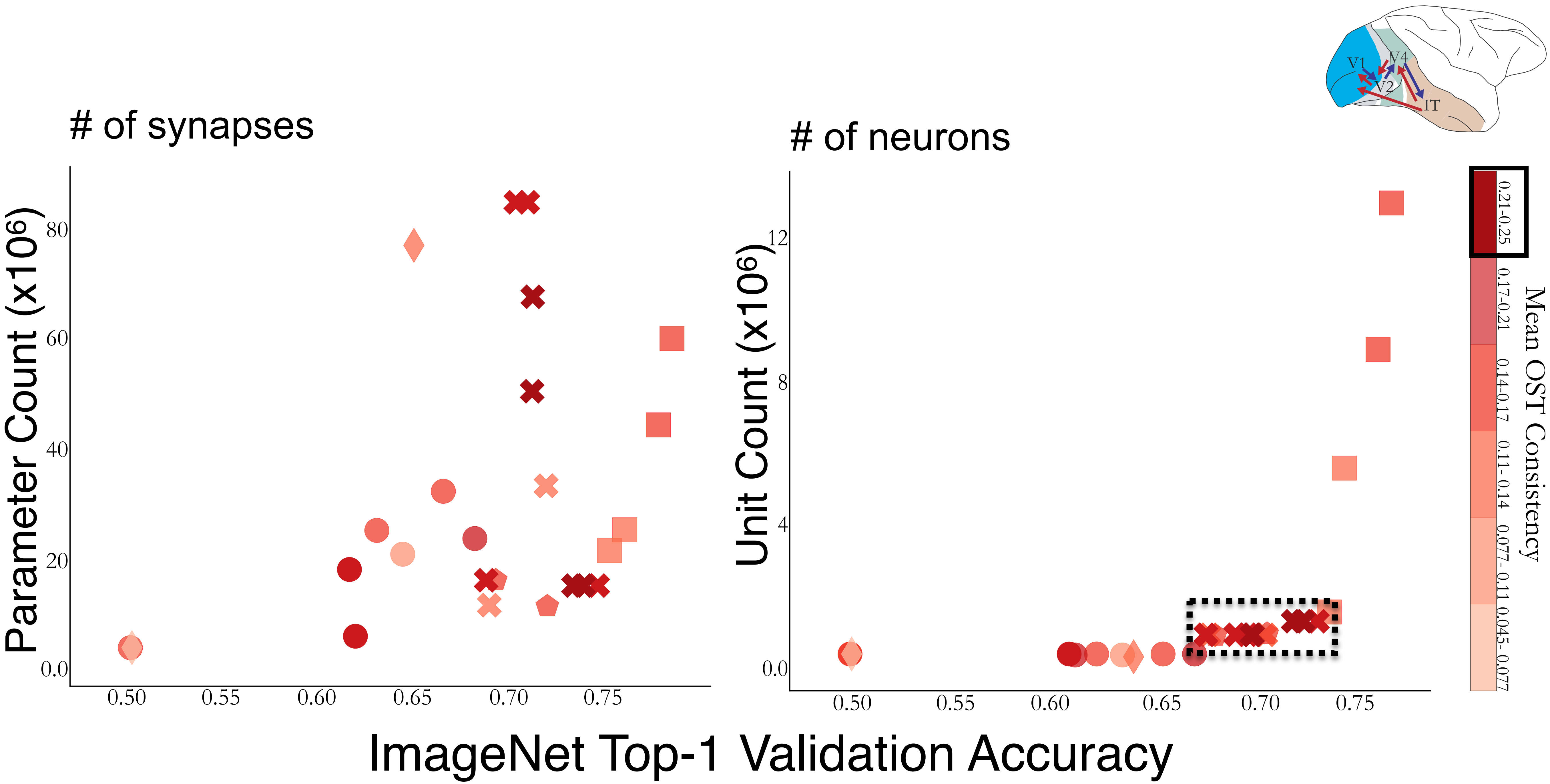
Why Are ConvRNNs the Most Brain-Like?



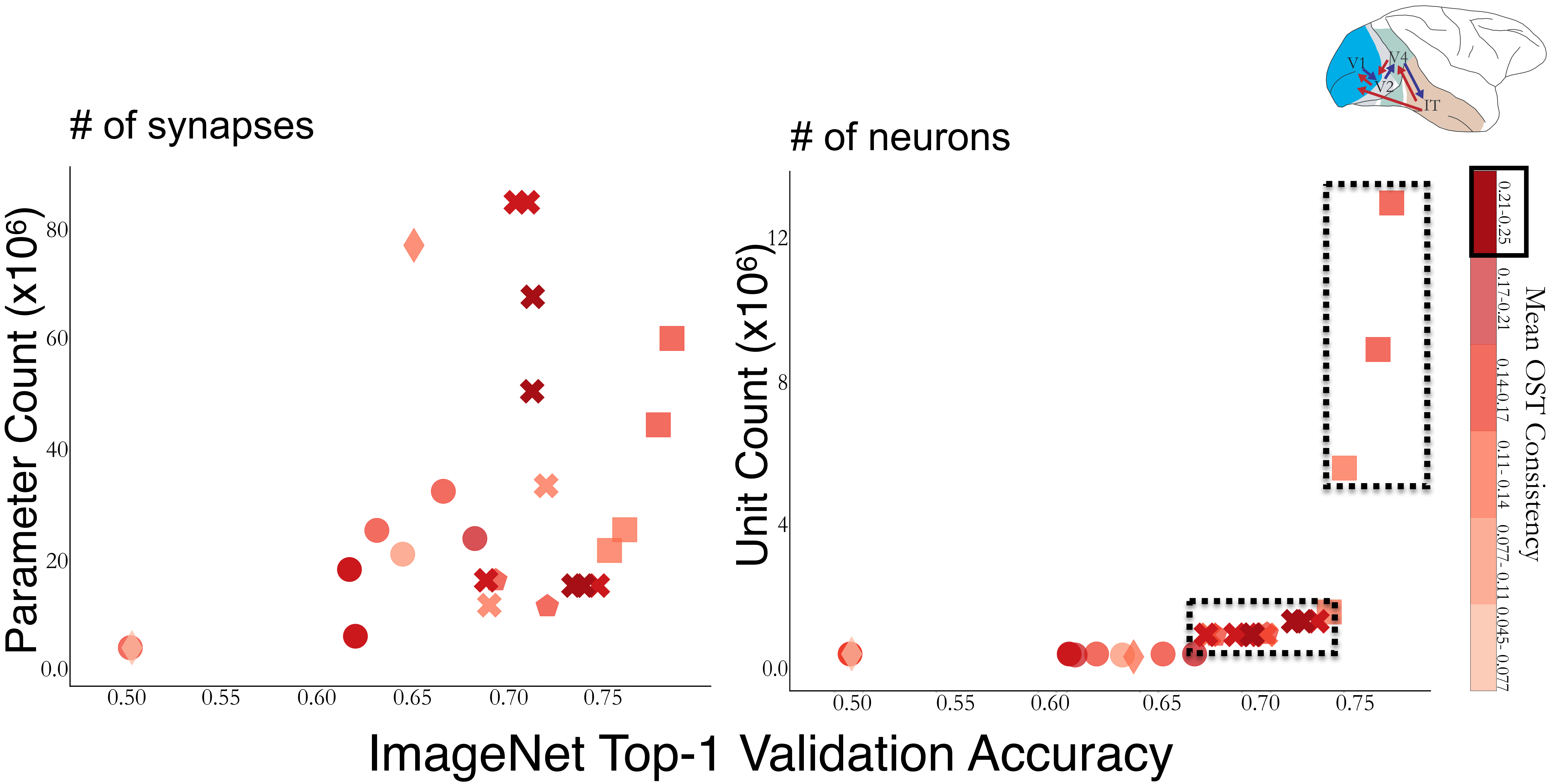
Why Are ConvRNNs the Most Brain-Like?



Smaller Networks that are Still *Performant* are More Brain-Like!



Smaller Networks that are Still *Performant* are More Brain-Like!



Outline

- Role of Recurrent Processing During Object Recognition

Enables more parameter/unit efficient models that gain object recognition performance by unrolling "deeper" in time, rather than adding more layers.

Moreso than simply "convolutionizing" standard LSTMs/GRUs.

- Recurrent Processing Best Explains Tactile Perception

Outline

- Role of Recurrent Processing During Object Recognition

Enables more parameter/unit efficient models that gain object recognition performance by unrolling "deeper" in time, rather than adding more layers.

Moreso than simply "convolutionizing" standard LSTMs/GRUs.

- Recurrent Processing Best Explains Tactile Perception

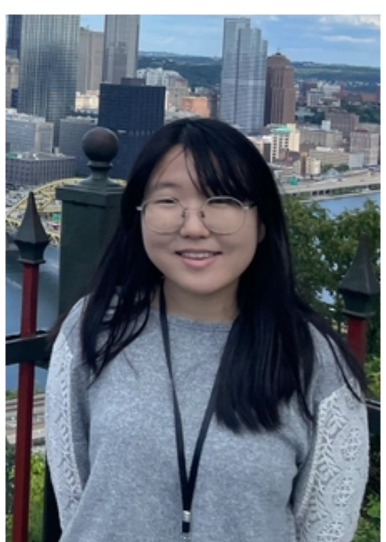
Tactile Processing

Task-Optimized Convolutional Recurrent Networks Align with Tactile Processing in the Rodent Brain

Trinity Chung[*,1], **Yuchen Shen**[*,2], **Nathan C. L. Kong**[4], **and Aran Nayebi**[2, 3, 1]

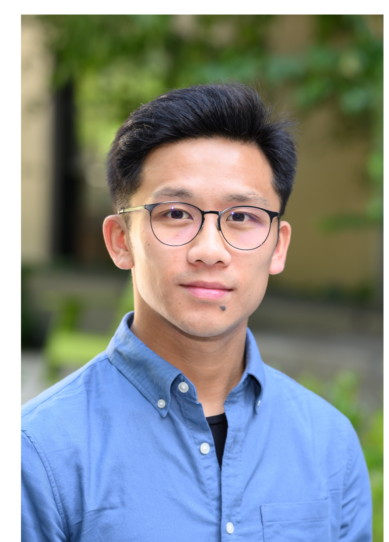
[1]Robotics Institute, Carnegie Mellon University; Pittsburgh, PA 15213
[2]Machine Learning Department, Carnegie Mellon University; Pittsburgh, PA 15213
[3]Neuroscience Institute, Carnegie Mellon University; Pittsburgh, PA 15213
[4]Department of Psychology, University of Pennsylvania; Philadelphia, PA 19104
[*] Equal contribution.
{trinityc, yuchens3, anayebi}@cs.cmu.edu; nclkong@sas.upenn.edu

NeurIPS 2025 Oral



Trinity Chung*

Yuchen Shen*

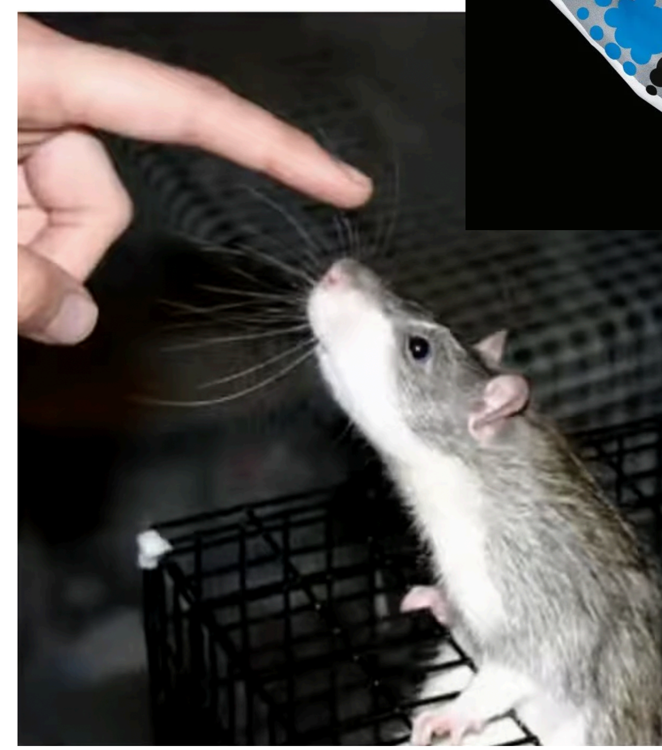


Nathan C.L. Kong

Motivation

- Tactile perception is still considerably under-explored in neuroscience and robotics
- Animals use whiskers, which are high temporal resolution tactile sensors

Q. What model architectures are brain-like for processing tactile data?



Methodology

Q. What model architectures are brain-like for processing tactile data?

1. Create whisking dataset
2. Search over model architectures
3. Evaluate on models on tactile recognition task
4. Compare models against real mouse brain data

Methodology

Q. What model architectures are brain-like for processing tactile data?

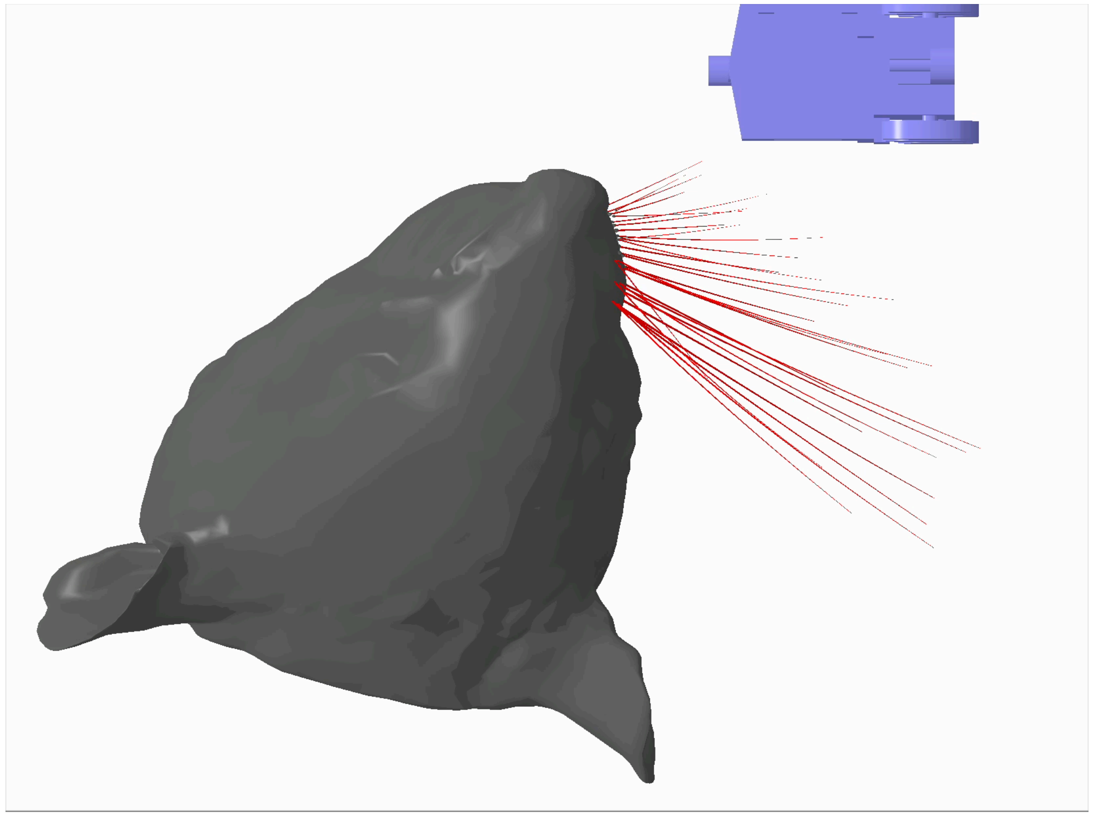
1. Create whisking dataset
2. Search over model architectures
3. Evaluate on models on tactile recognition task
4. Compare models against real mouse brain data

Methodology

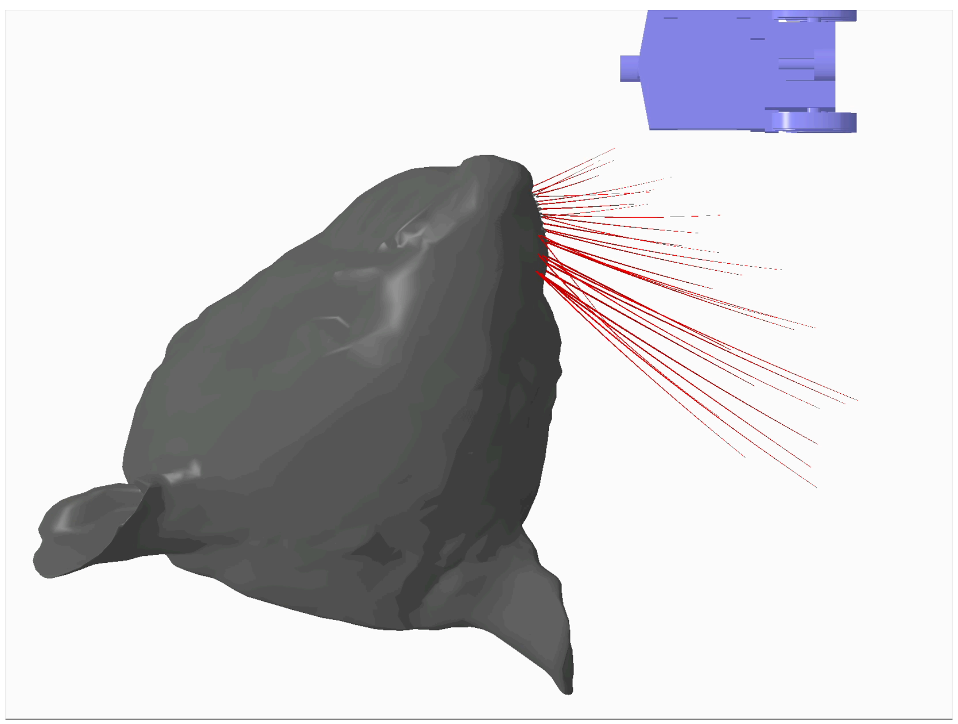
Q. What model architectures are brain-like for processing tactile data?

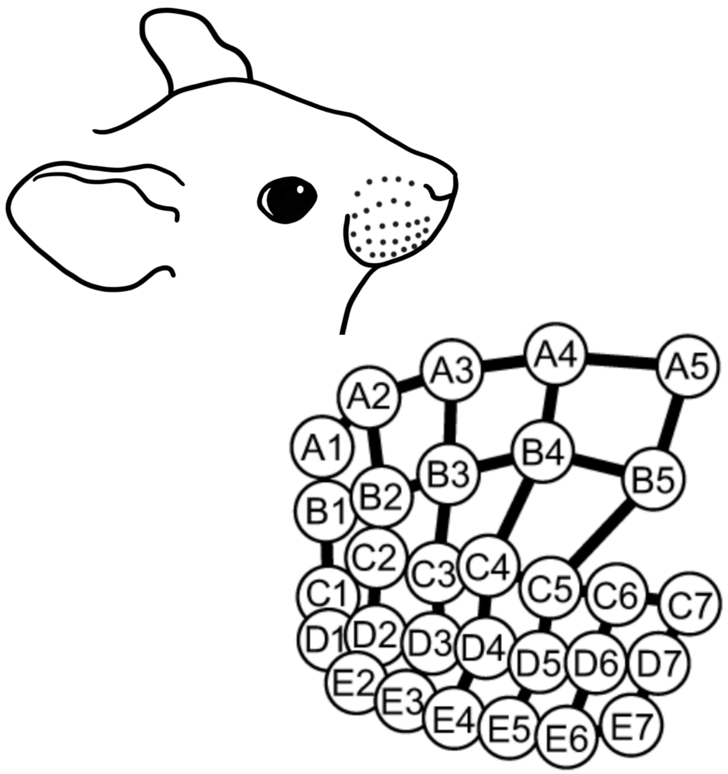
1. **Create whisking dataset**
2. Search over model architectures
3. Evaluate on models on tactile recognition task
4. Compare models against real mouse brain data

Dataset/Task: **Categorizing Shapes by Whisking**

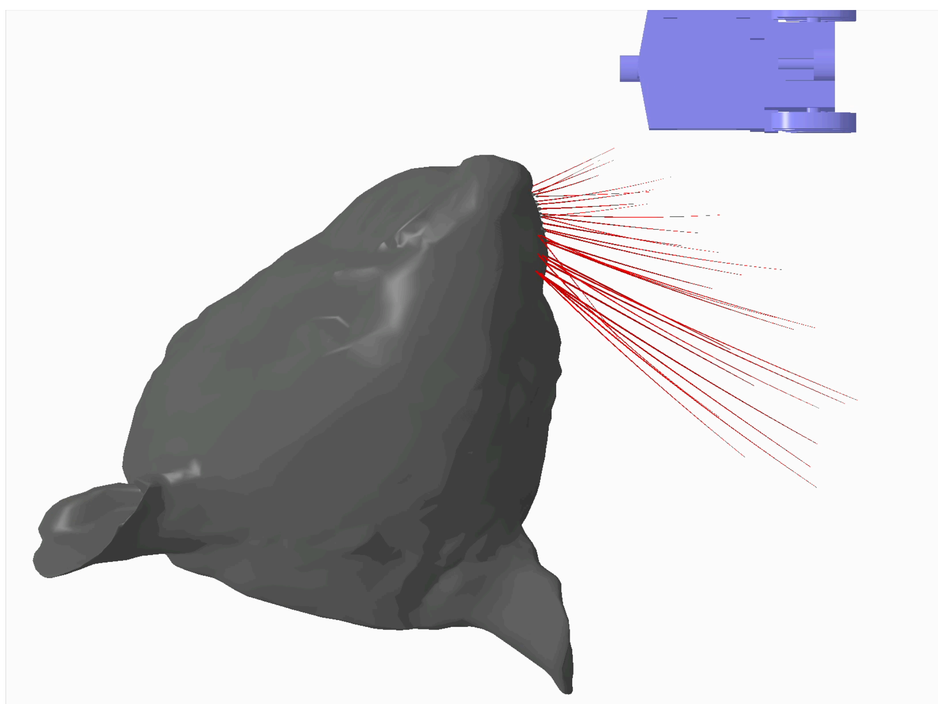


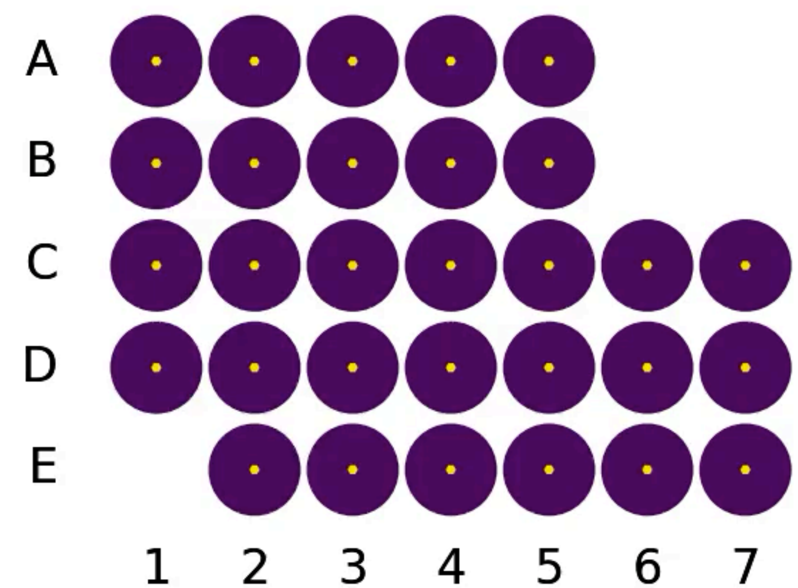
Dataset/Task: **Categorizing Shapes by Whisking**





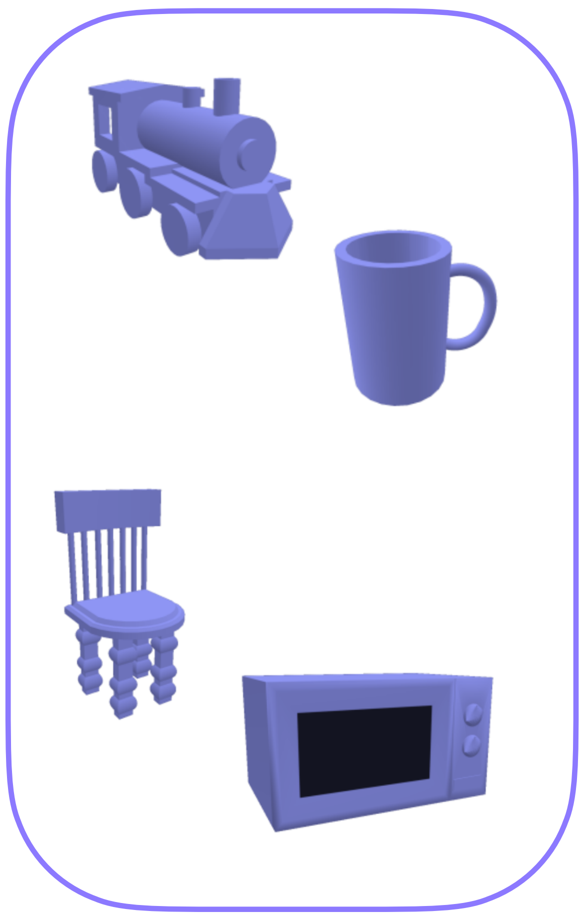
Dataset/Task: **Categorizing Shapes by Whisking**



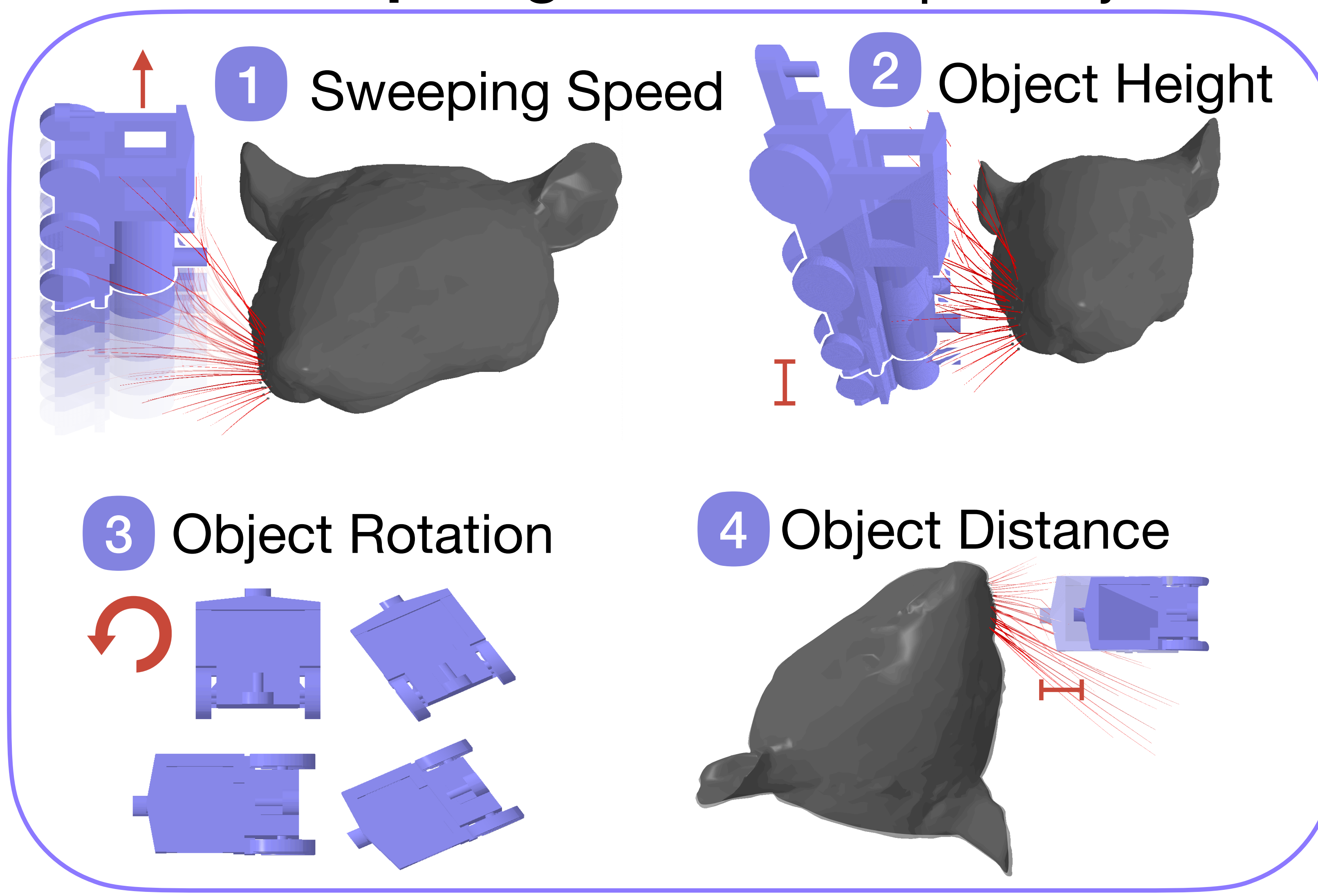


Dataset/Task: **Categorizing Shapes by Whisking**

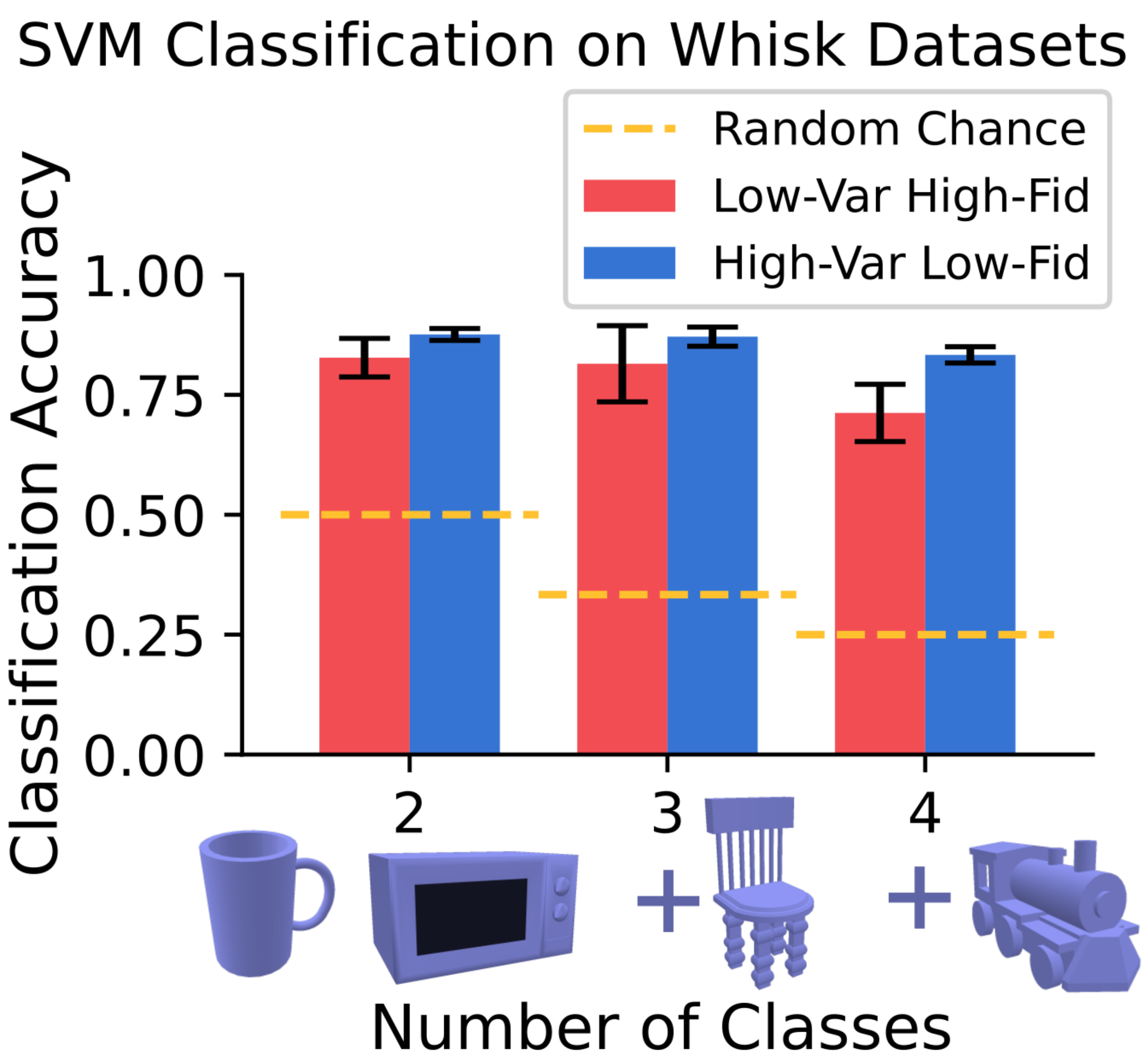
9881 Objects
from **ShapeNet**



Sweep Augmentations per object



Dataset/Task: **Categorizing Shapes by Whisking**



Methodology

Q. What model architectures are
brain-like for processing tactile data?

1. **Create whisking dataset**
2. Search over model architectures
3. Evaluate on models on tactile recognition task
4. Compare models against real mouse brain data

Methodology

Q. What model architectures are brain-like for processing tactile data?

1. Create whisking dataset
2. **Search over model architectures**
3. Evaluate on models on tactile recognition task
4. Compare models against real mouse brain data

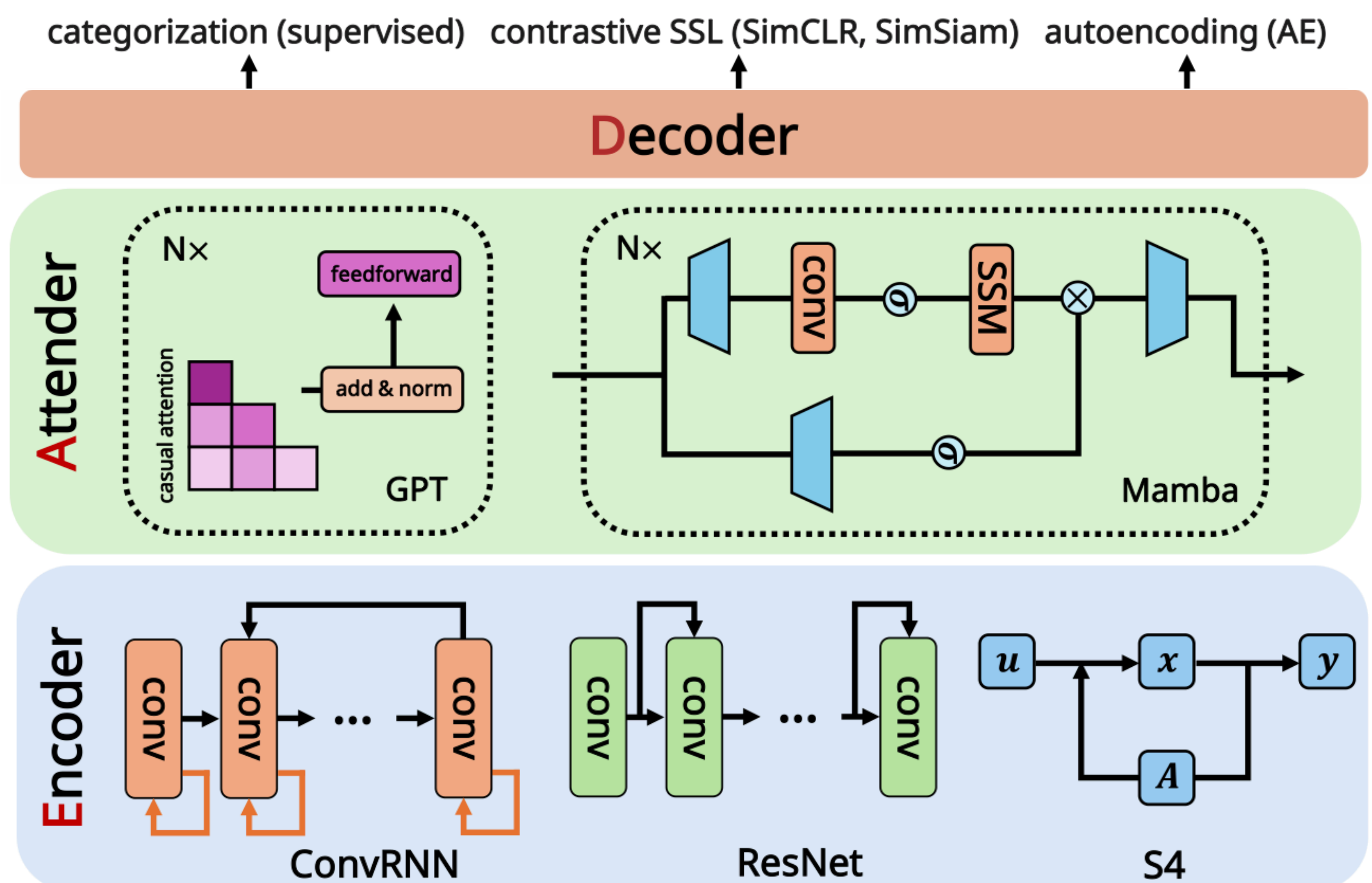
Searching through Model Architectures

- Construct temporal model architectures using our **Encoder-Attender-Decoder (EAD)** framework

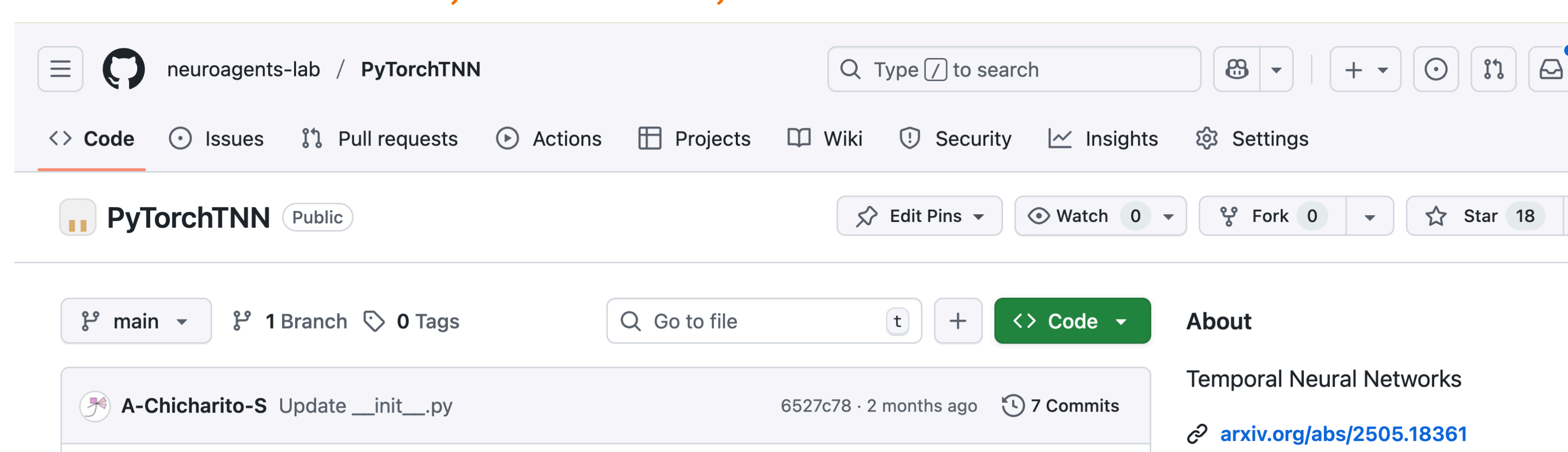
- Update rule for ConvRNNs

$$h_t^\ell = C_\ell\left(F_\ell\left(\bigoplus_{j\neq\ell} r_t^j\right), h_{t-1}^\ell\right)$$
$$r_t^\ell = A_\ell(h_t^\ell),$$

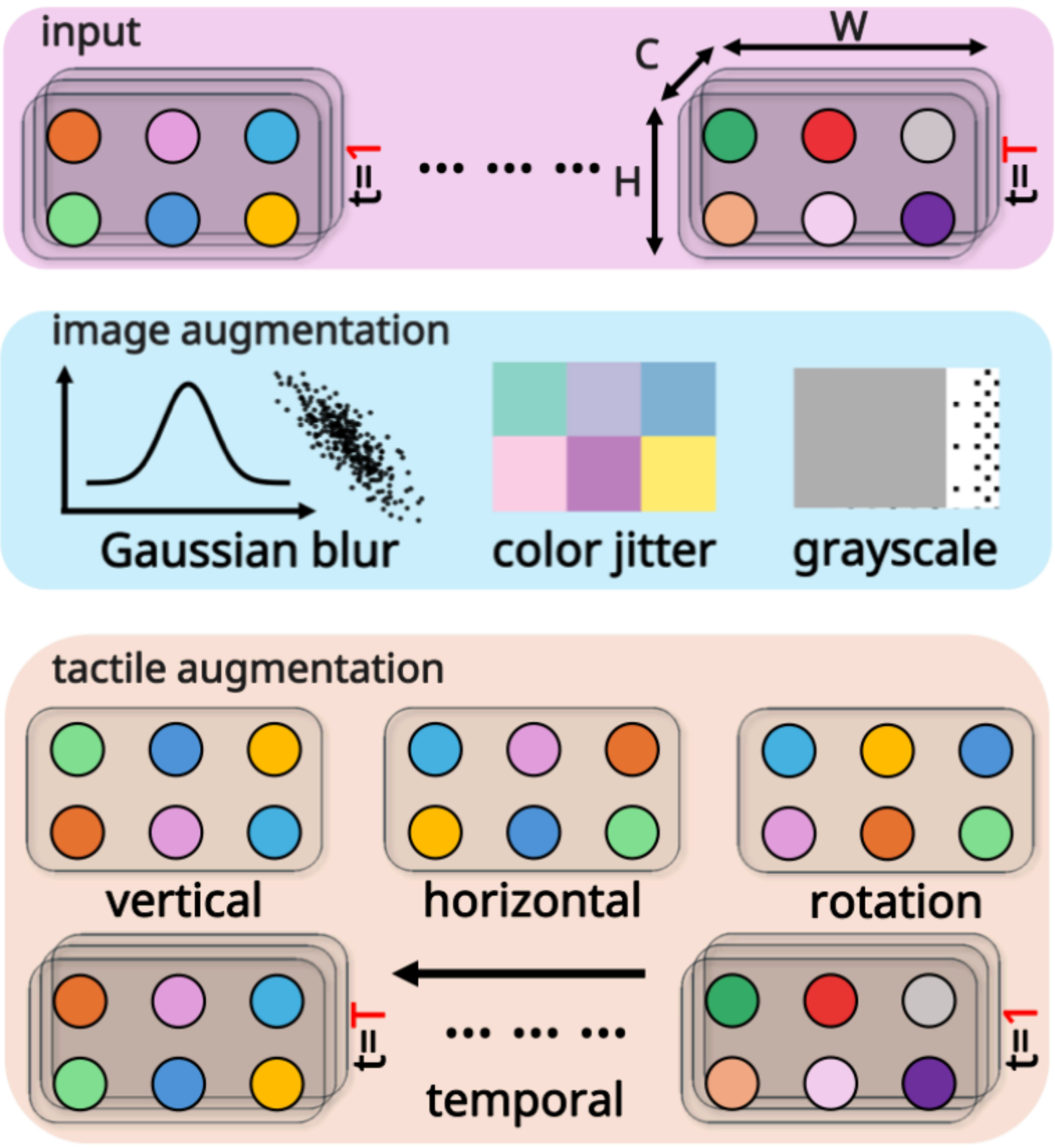
- github.com/neuroagents-lab/ [PyTorchTNN](https://github.com/neuroagents-lab/PyTorchTNN) library



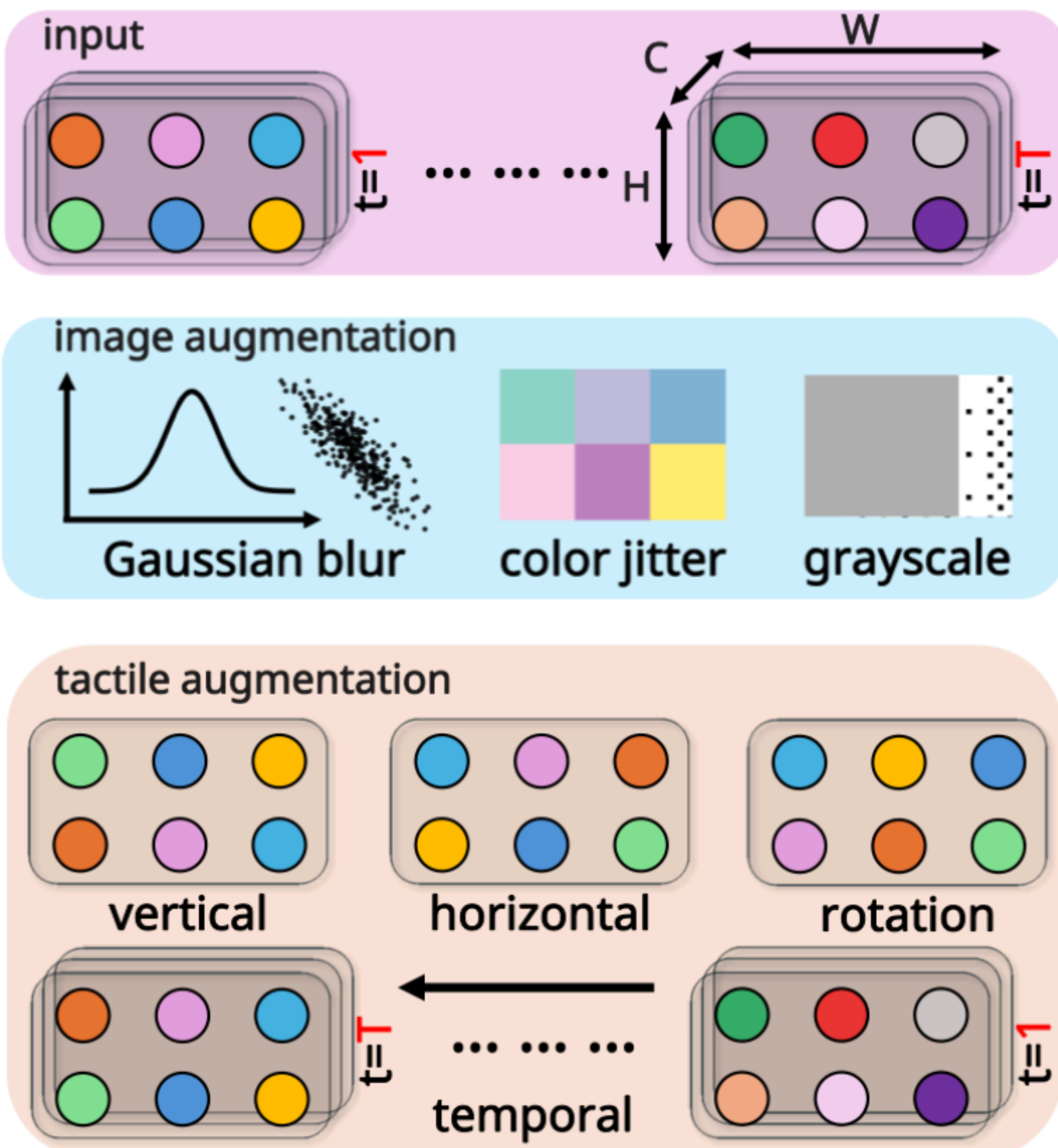
InterRNN, UGRNN, ...



Training with **Image vs Tactile Augmentations**



Training with **Image vs Tactile Augmentations**



Methodology

Q. What model architectures are brain-like for processing tactile data?

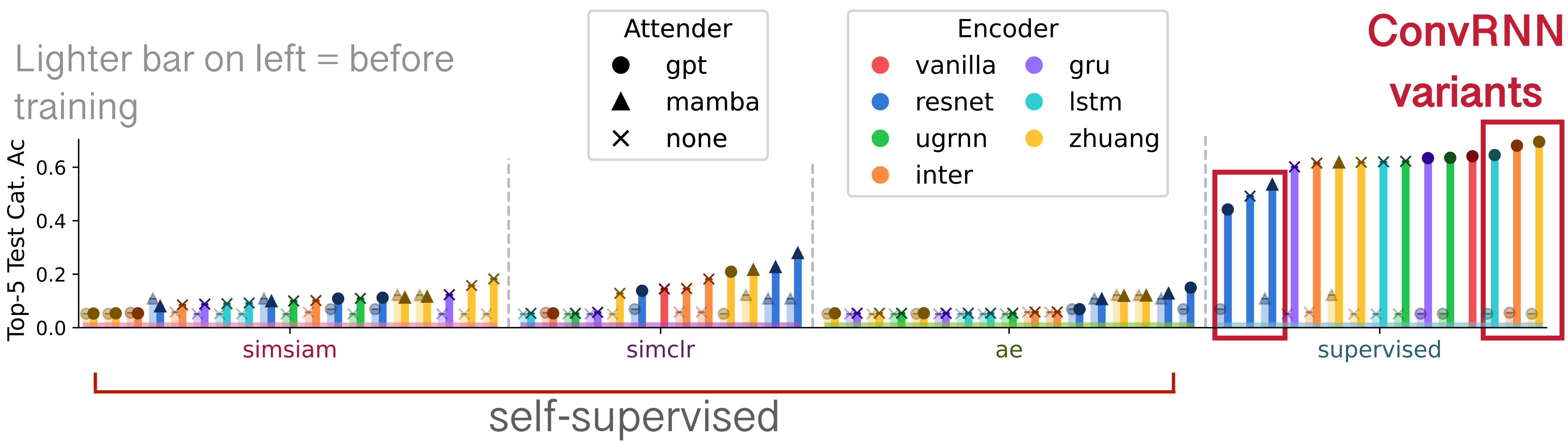
1. Create whisking dataset
2. **Search over model architectures**
3. Evaluate on models on tactile recognition task
4. Compare models against real mouse brain data

Methodology

Q. What model architectures are brain-like for processing tactile data?

1. Create whisking dataset
2. Search over model architectures
3. **Evaluate on models on tactile recognition task**
4. Compare models against real mouse brain data

Results on Task Performance



ConvRNNs outperform non-recurrent architectures on the tactile recognition task.

Methodology

Q. What model architectures are brain-like for processing tactile data?

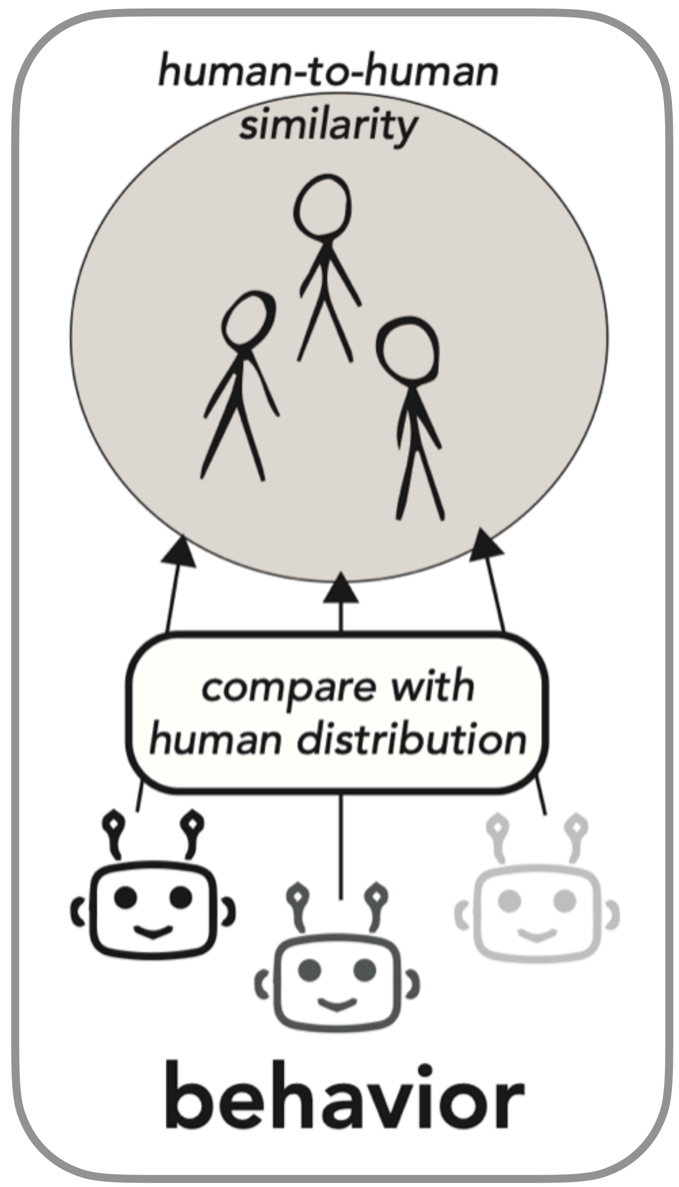
1. Create whisking dataset
2. Search over model architectures
3. **Evaluate on models on tactile recognition task**
4. Compare models against real mouse brain data

Methodology

Q. What model architectures are brain-like for processing tactile data?

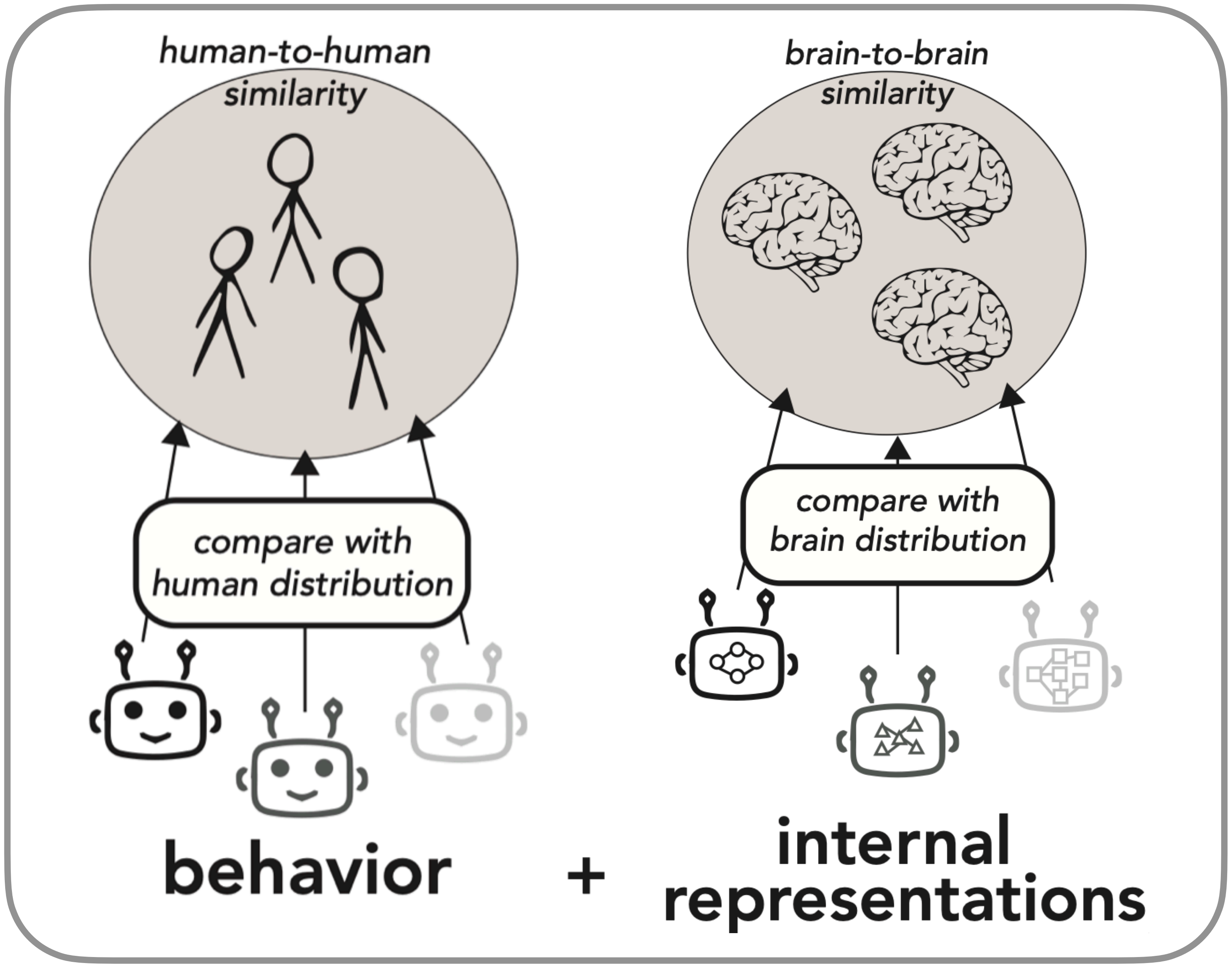
1. Create whisking dataset
2. Search over model architectures
3. Evaluate on models on tactile recognition task
4. **Compare models against real mouse brain data**

Turing Test



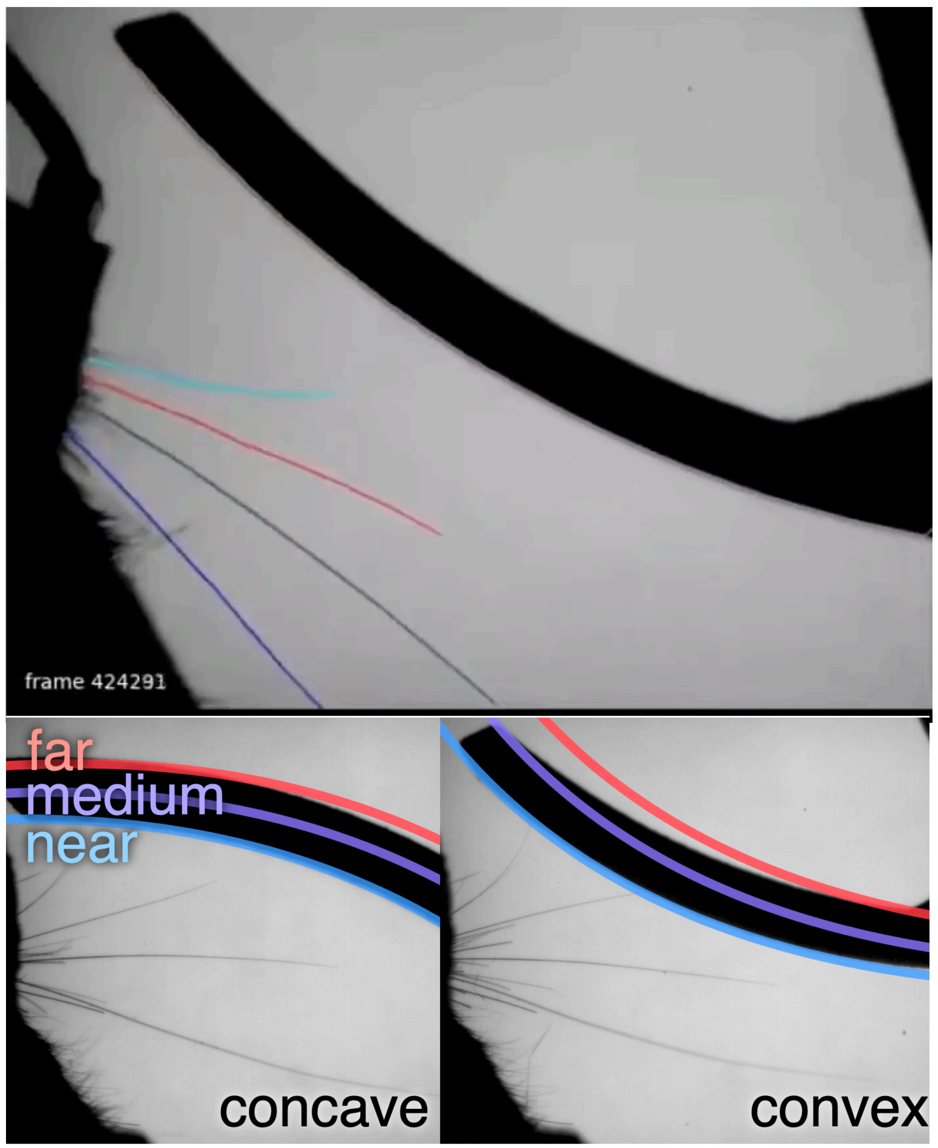
"Brain-Model Evaluations Need the NeuroAI Turing Test," Feather* et al. 2025

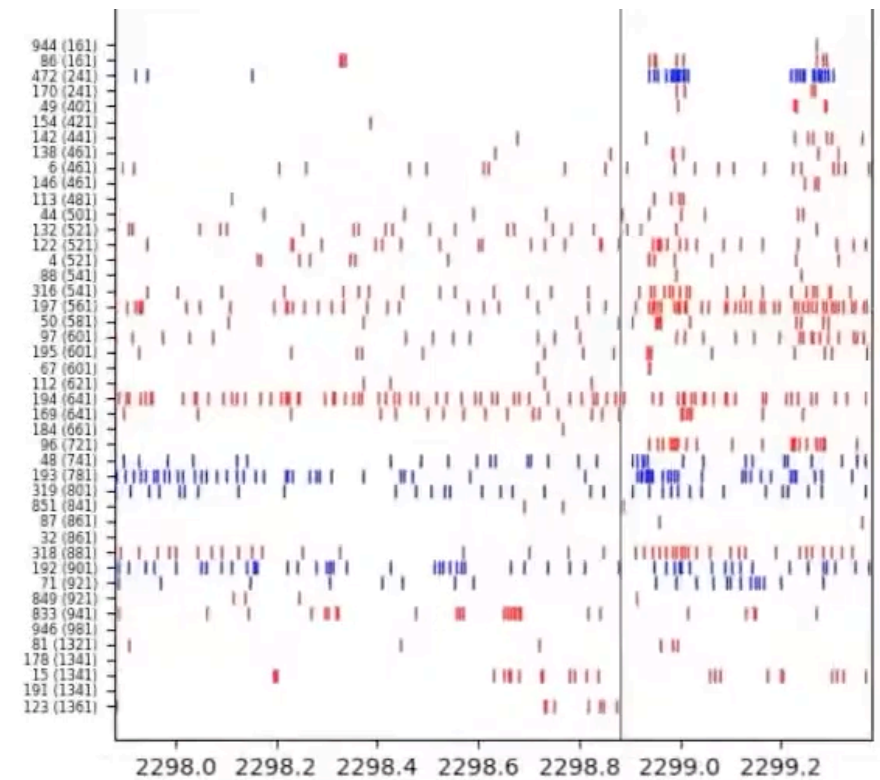
NeuroAI Turing Test



"Brain-Model Evaluations Need the NeuroAI Turing Test," Feather* et al. 2025

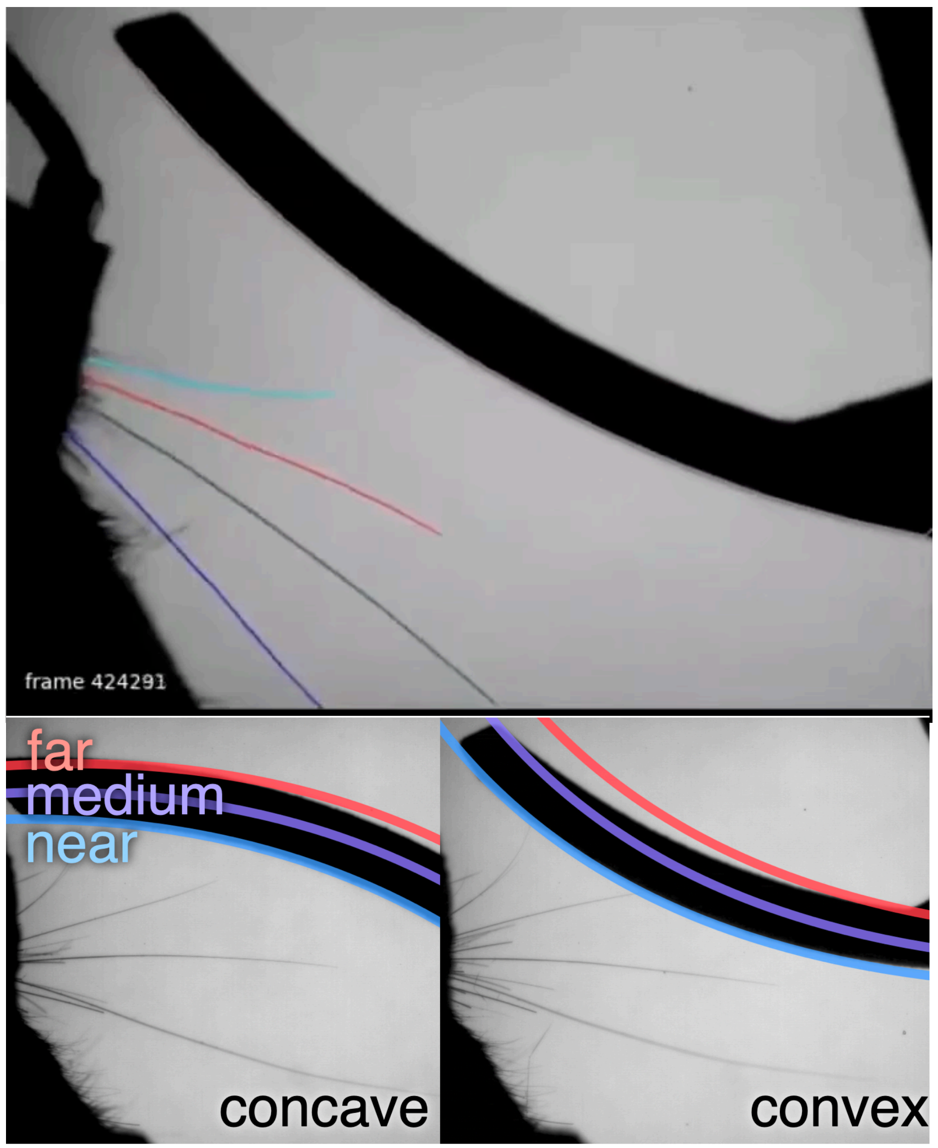
Neural Data


frame 424291
far
medium
near
concave
convex

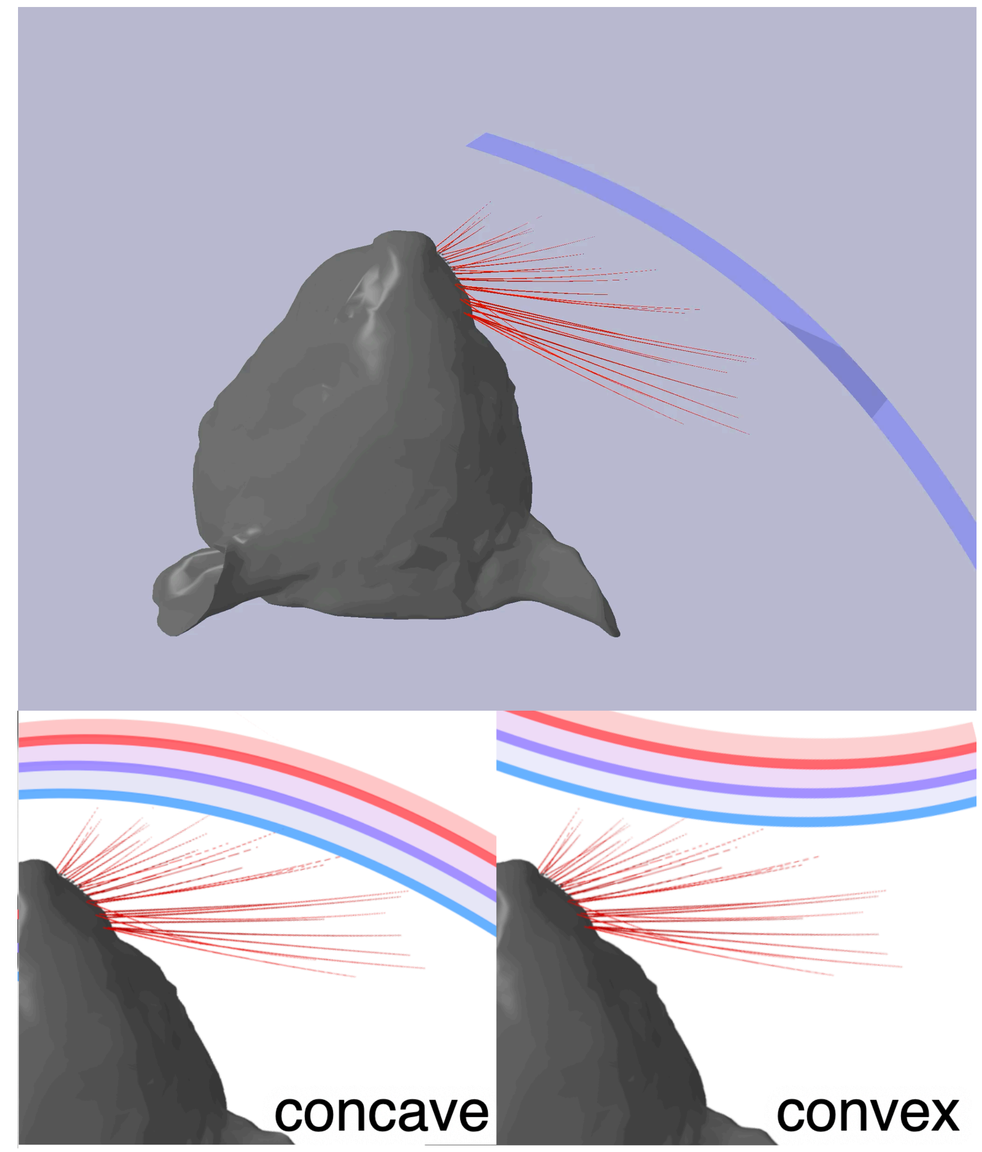

2298.0
2298.2
2298.4
2298.6
2298.8
2299.0
2299.2

Neural Data

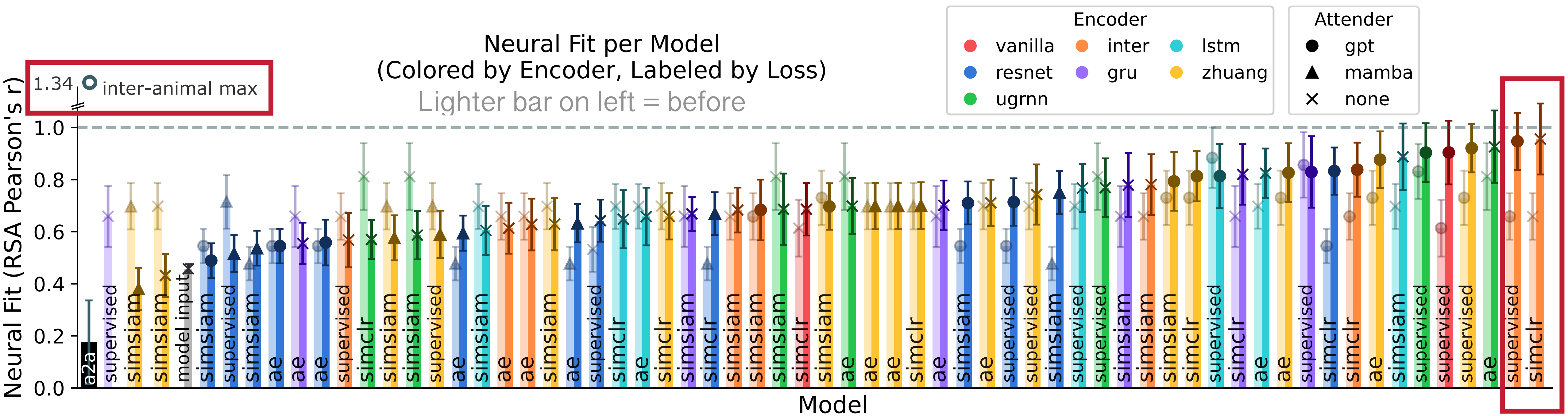
Real

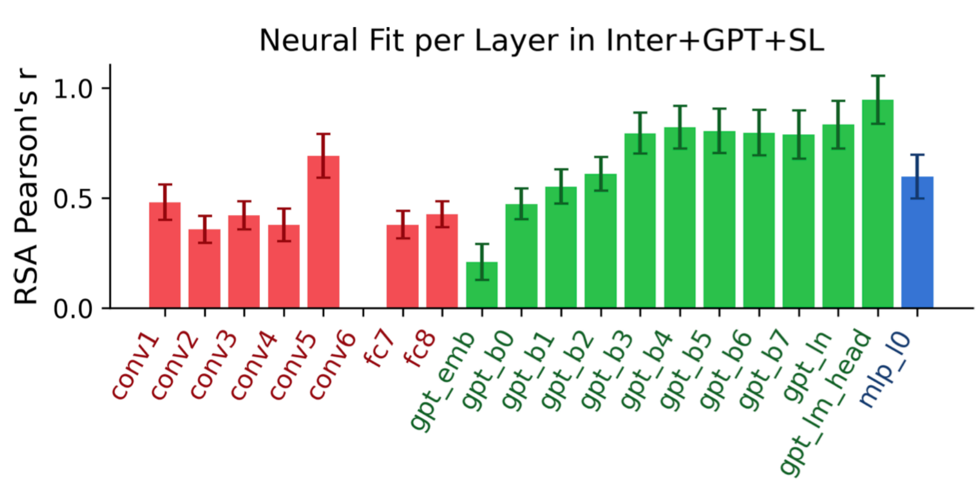


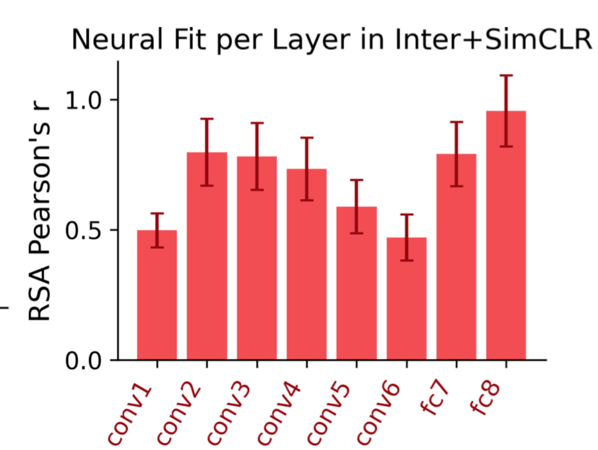
Sim



Results on Neural Evaluation







Update rules for **IntersectionRNN**

$$m_t^\ell = \tanh\left(W_m^\ell * x_t^\ell + U_m^\ell * s_{t-1}^\ell + b_m^\ell\right)$$
$$n_t^\ell = \mathrm{ReLU}\left(W_n^\ell * x_t^\ell + U_n^\ell * s_{t-1}^\ell + b_n^\ell\right)$$
$$p_t^\ell = \sigma\left(W_p^\ell * x_t^\ell + U_p^\ell * s_{t-1}^\ell + b_p^\ell + 1\right)$$
$$y_t^\ell = \sigma\left(W_y^\ell * x_t^\ell + U_y^\ell * s_{t-1}^\ell + b_y^\ell + 1\right)$$
$$s_t^\ell = p_t^\ell \circ s_{t-1}^\ell + \left(1 - p_t^\ell\right) \circ m_t^\ell$$
$$h_t^\ell = y_t^\ell \circ x_t^\ell + \left(1 - y_t^\ell\right) \circ n_t^\ell$$

Methodology

Q. What model architectures are
brain-like for processing tactile forces?

1. Create whisking dataset
2. Search over model architectures
3. Evaluate on models on tactile recognition task
4. **Compare models against real mouse brain data**

Methodology

Q. What model architectures are brain-like for processing tactile forces?

1. Create whisking dataset
2. Search over model architectures
3. Evaluate on models on tactile recognition task
4. Compare models against real mouse brain data

Conclusions

Q. What model architectures are brain-like for processing tactile forces?

ConvRNNs!

Conclusions

Q. What model architectures are brain-like for processing tactile forces?

ConvRNNs!

- **ConvRNNs outperform** feedforward/SSMs on realistic tactile recognition and best match neural responses in mouse barrel cortex.
- **Contrastive SSL** matches supervised neural alignment, possibly suggesting a **general-purpose representation** in the somatosensory cortex.

Outline

- Role of Recurrent Processing During Object Recognition

Enables more parameter/unit efficient models that gain object recognition performance by unrolling "deeper" in time, rather than adding more layers.

Moreso than simply "convolutionizing" standard LSTMs/GRUs.

- Recurrent Processing Best Explains Tactile Perception

Here ConvRNNs actually seem to help with the task, not just saving on space. Perhaps because it is inherently more temporal?

Shared principles across sensory modalities?

Limitations & Future Work

- Reevaluate on neural datasets with more stimuli
- Explore types of tactile sensors and the most useful form of tactile data for learning manipulation tasks
- Multimodal fusion methods
  - Vision, proprioception
  - Concatenation, attention, routing

Agents?

Intrinsic Goals for Autonomous Agents: Model-Based Exploration in Virtual Zebrafish Predicts Ethological Behavior and Whole-Brain Dynamics

Reece Keller[1,2*] **Alyn Kirsch** [2] **Felix Pei**[1] **Xaq Pitkow**[1,3]
Leo Kozachkov[4,†] **Aran Nayebi**[3,1,2,†]

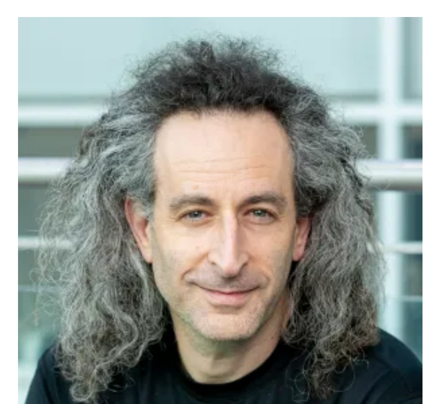
First autonomous agent that can predict *whole-brain* data!

NeurIPS 2025

Reece Keller

Alyn Tornell

Felix Pei

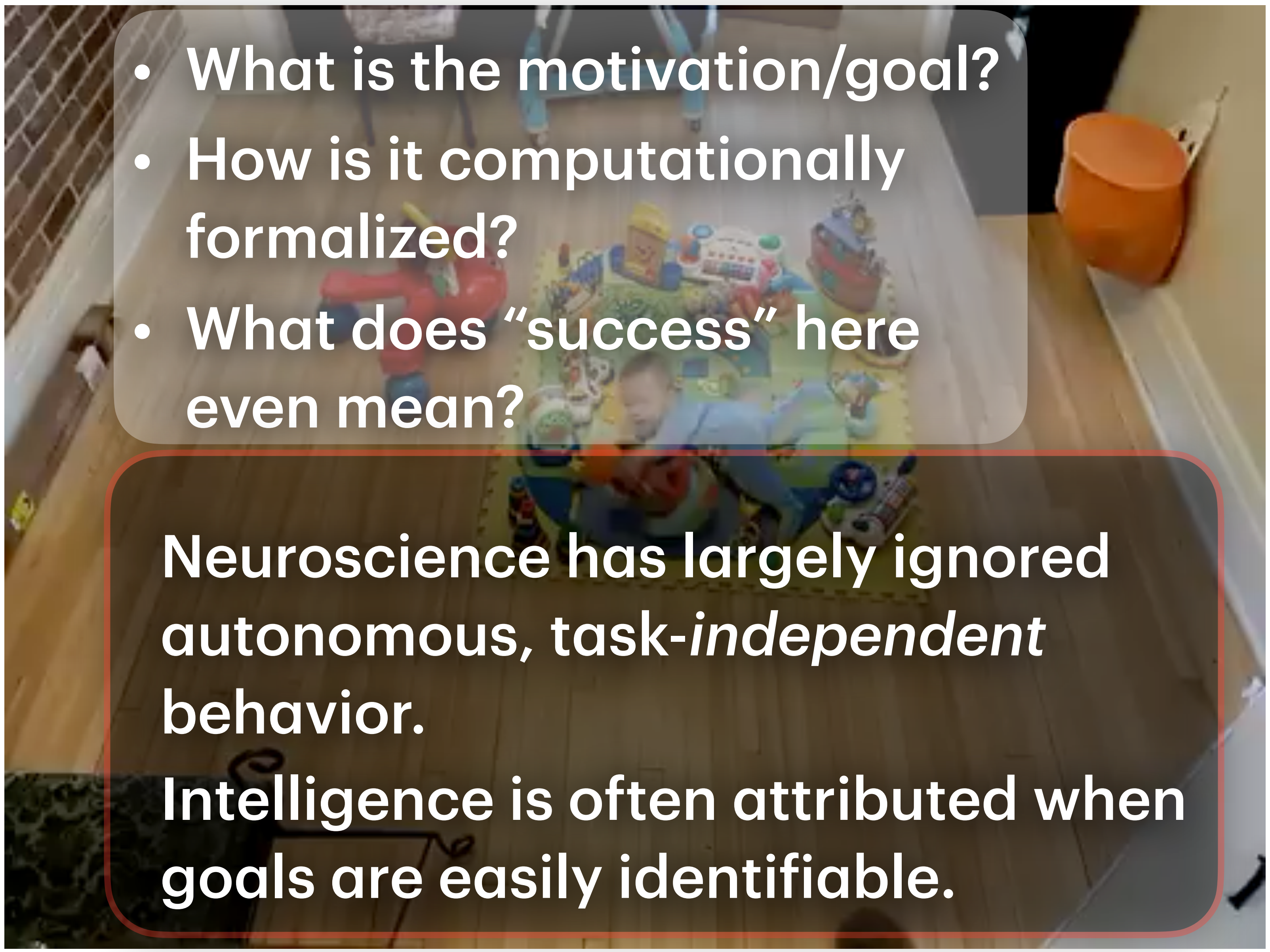


Xaq Pitkow

Leo Kozachkov†

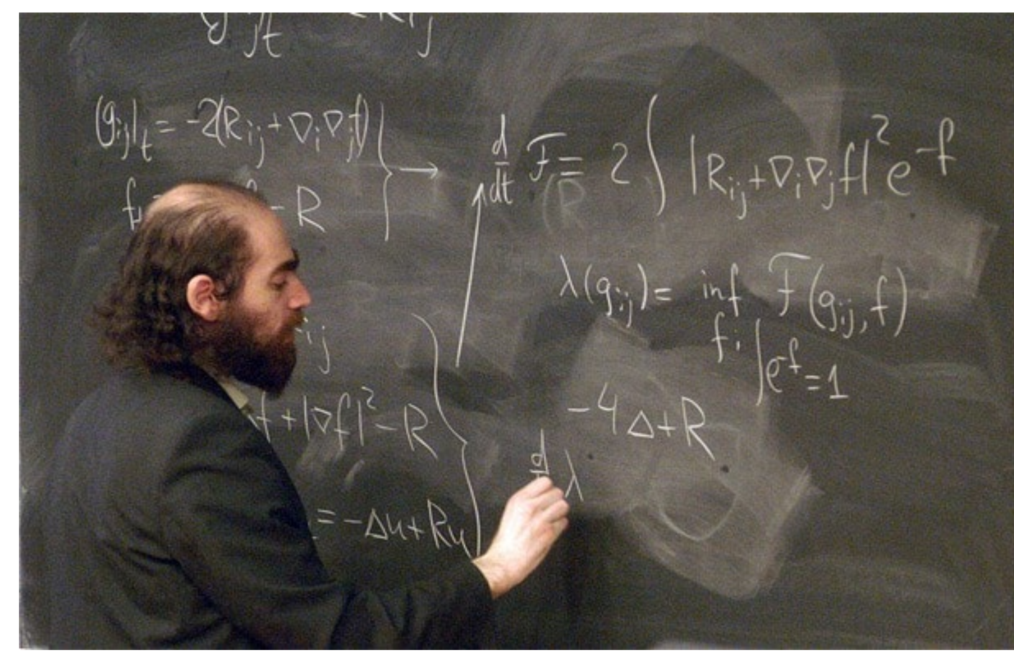
Why is Animal Autonomy Hard?

The behavioral repertoire is enormous...



- **What is the motivation/goal?**
- **How is it computationally formalized?**
- **What does "success" here even mean?**

Neuroscience has largely ignored autonomous, task-*independent* behavior.
Intelligence is often attributed when goals are easily identifiable.



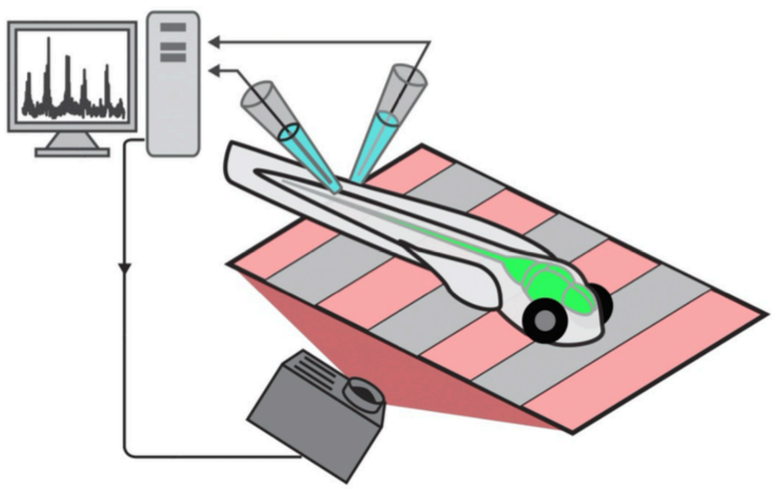
Unlike games where RL has succeeded, the environment doesn't have a dense reward function. It must be (somehow) *internally* generated by the organism!

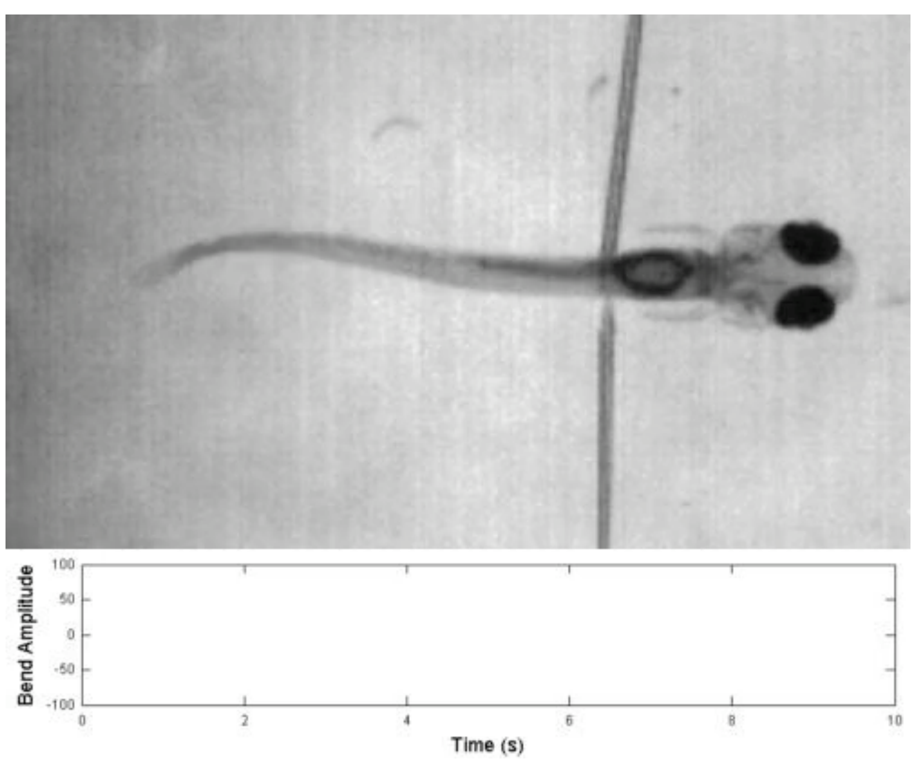
Slides credit: Reece Keller

Glia Accumulate Evidence that Actions Are Futile and Suppress Unsuccessful Behavior

Yu Mu,[1,4,*] Davis V. Bennett,[1,2,4] Mikail Rubinov,[1,3,4] Sujatha Narayan,[1] Chao-Tsung Yang,[1] Masashi Tanimoto,[1] Brett D. Mensh,[1] Loren L. Looger,[1] and Misha B. Ahrens[1,5,*]

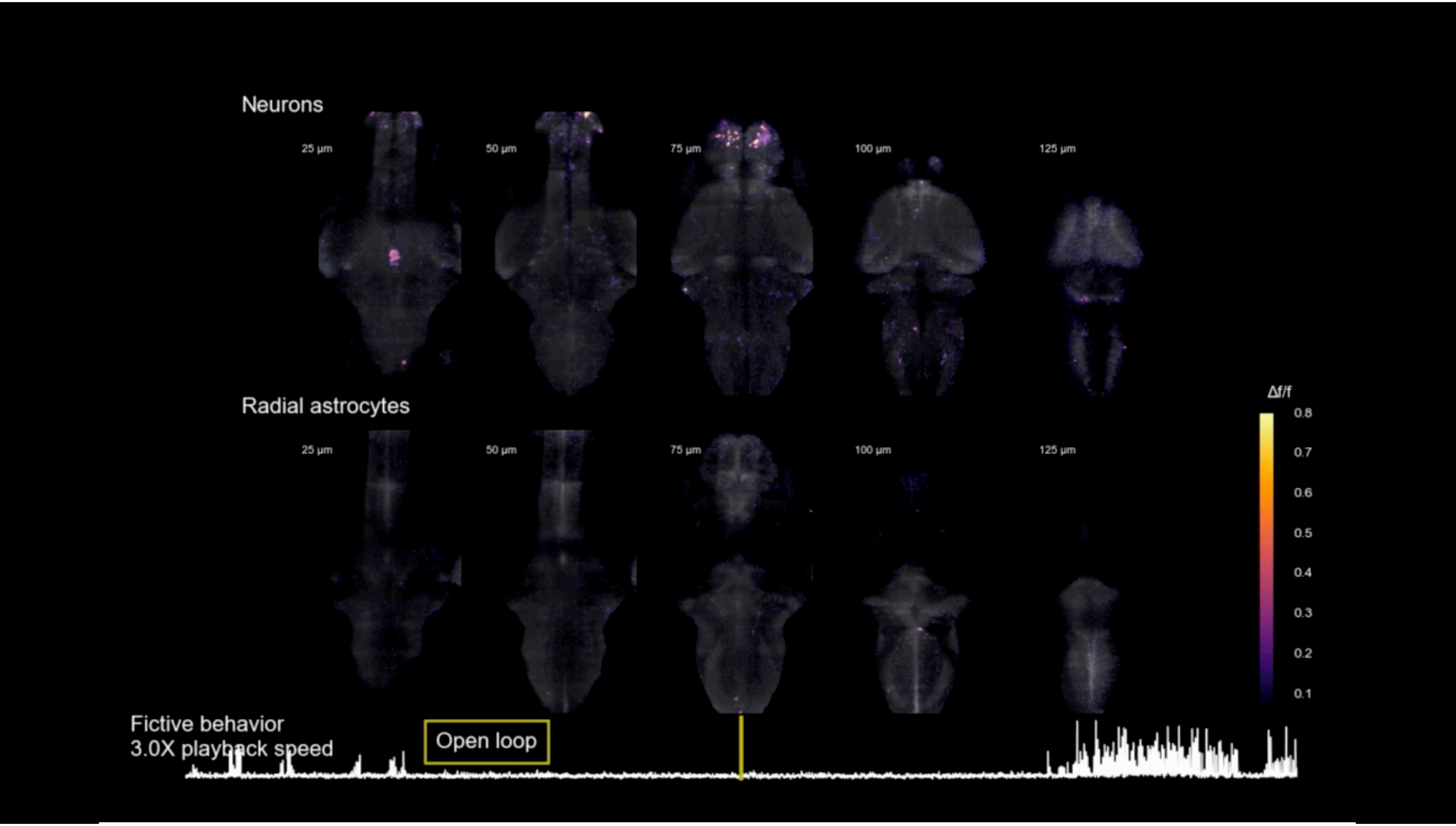
virtual reality navigation





1. Ecologically-relevant environment
2. "Cognitive" states with clear behavioral readouts
3. Large-scale multi-area neural recordings

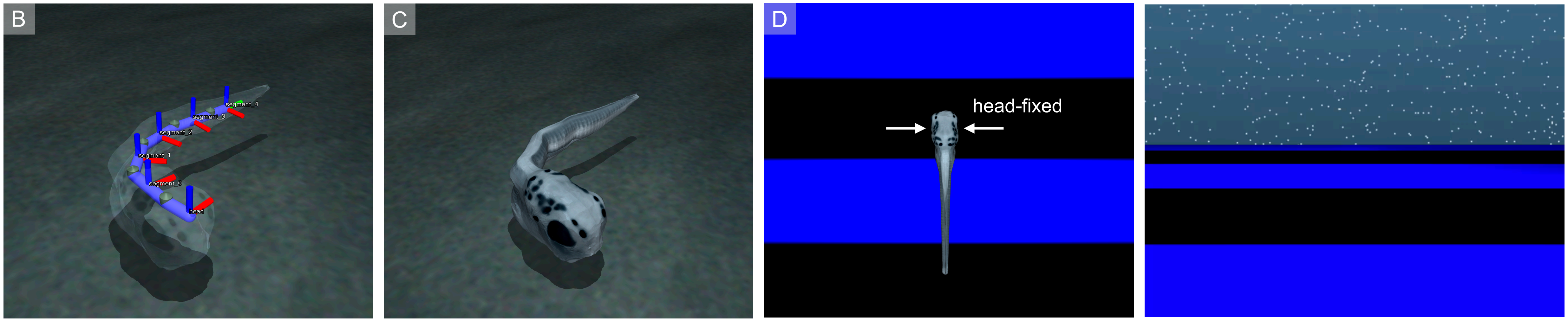
Mu et al., *Cell* (2019)


Neurons
25 µm
50 µm
75 µm
100 µm
125 µm
Radial astrocytes
25 µm
50 µm
75 µm
100 µm
125 µm
Δf/f
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Fictive behavior
3.0X playback speed
Open loop

Question: What intrinsic drive explains this behavior?

Specifically, how should world-models be used to guide autonomous decisions in real-world situations (e.g. encountering unseen physics)?

Zebrafish Simulation Environment



Actuation

- The embodiment must afford a faithful comparison with the animal behavior.
- Behavioral signal is low dimensional -> embodiment can be low dimensional
- Open-source embodiments that capture basic ethology already exist!

Sensing

- The zebrafish behavior is driven by optic flow and proprioception. A basic vision model and state information is sufficient.

Our philosophy:

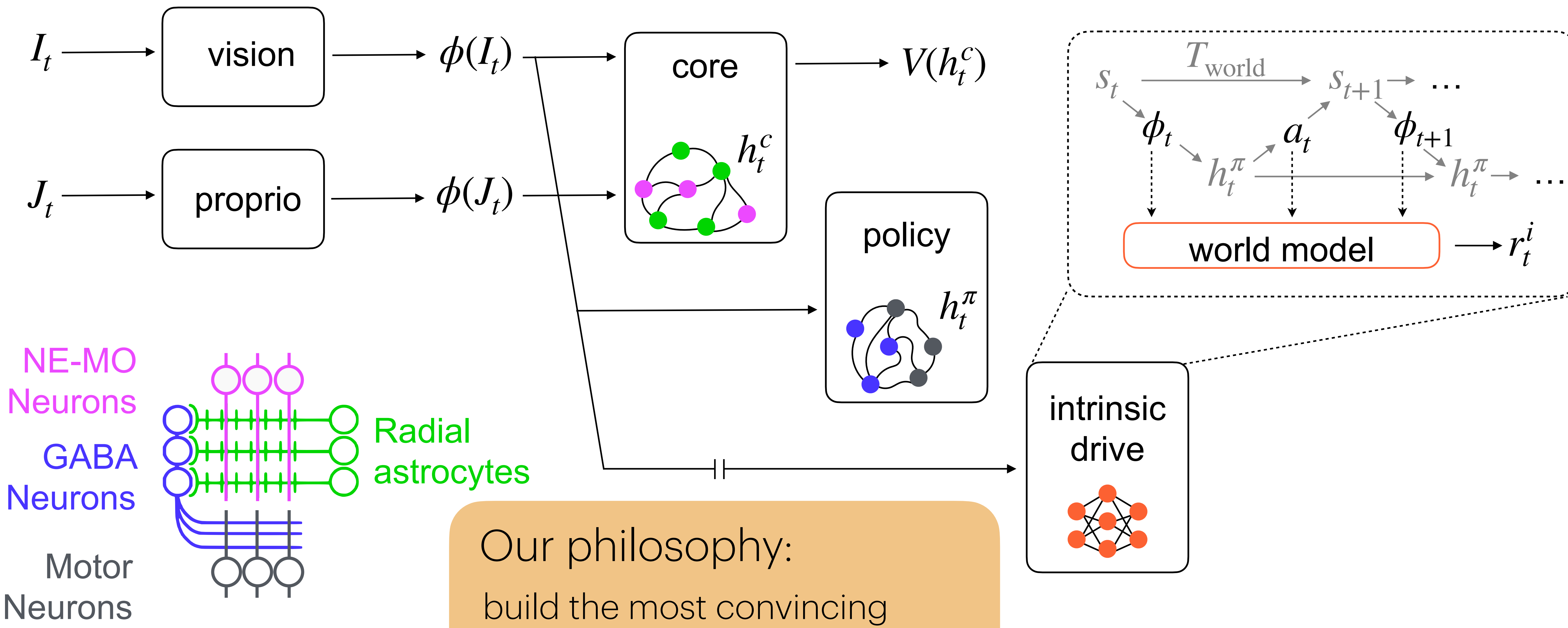
build the most convincing model possible.

- stimulus/image computable
- realistic physics
- flexible parameterization

Question: What intrinsic drive explains this behavior?

Specifically, how should world-models be used to guide autonomous decisions in real-world situations (e.g. encountering unseen physics)?

Zebrafish Agent Architecture



Our philosophy:

build the most convincing model possible.

- stimulus/image computable
- realistic physics
- flexible parameterization

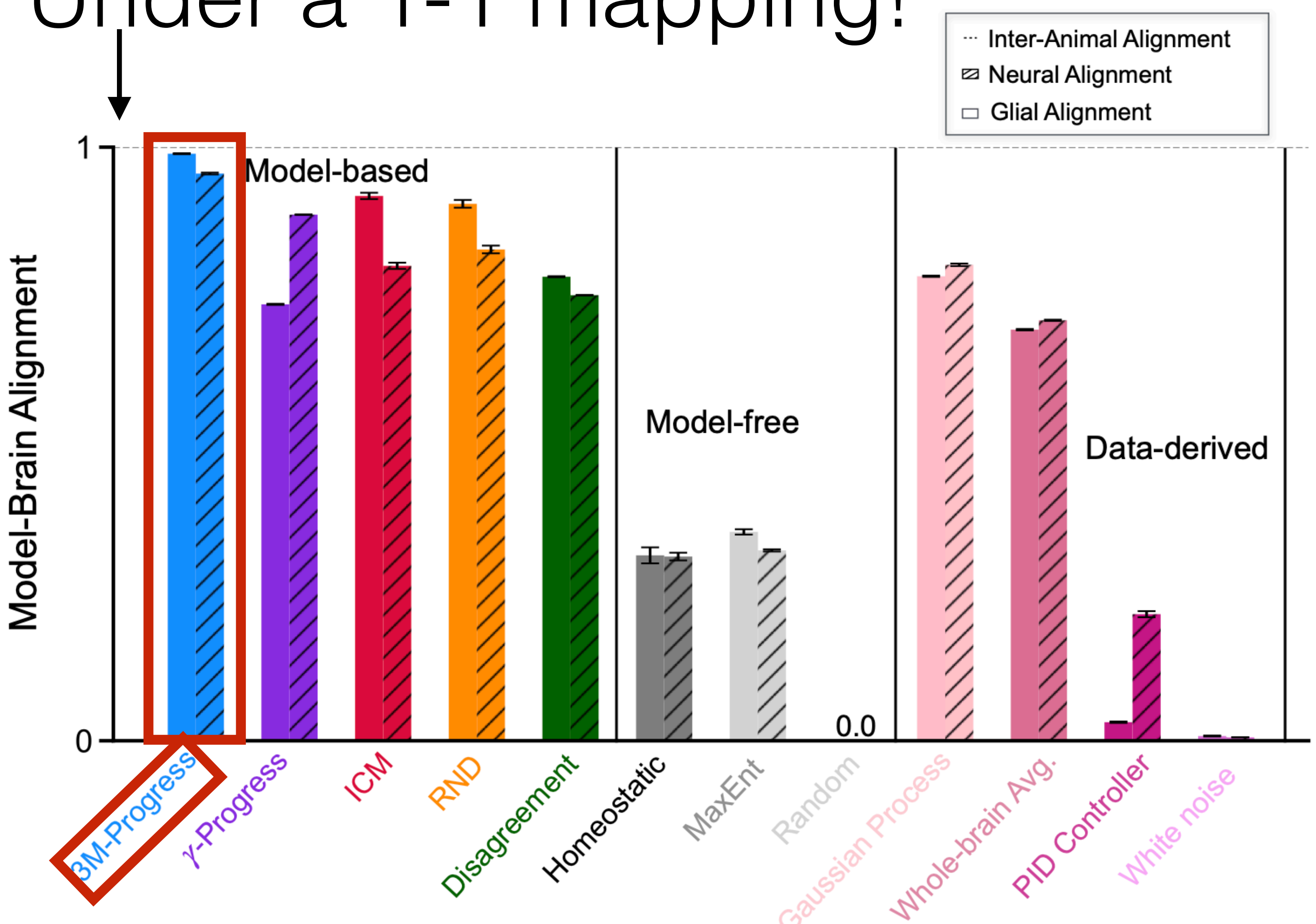
Putting it all together

3M-Progress Captures Whole-Brain Dynamics

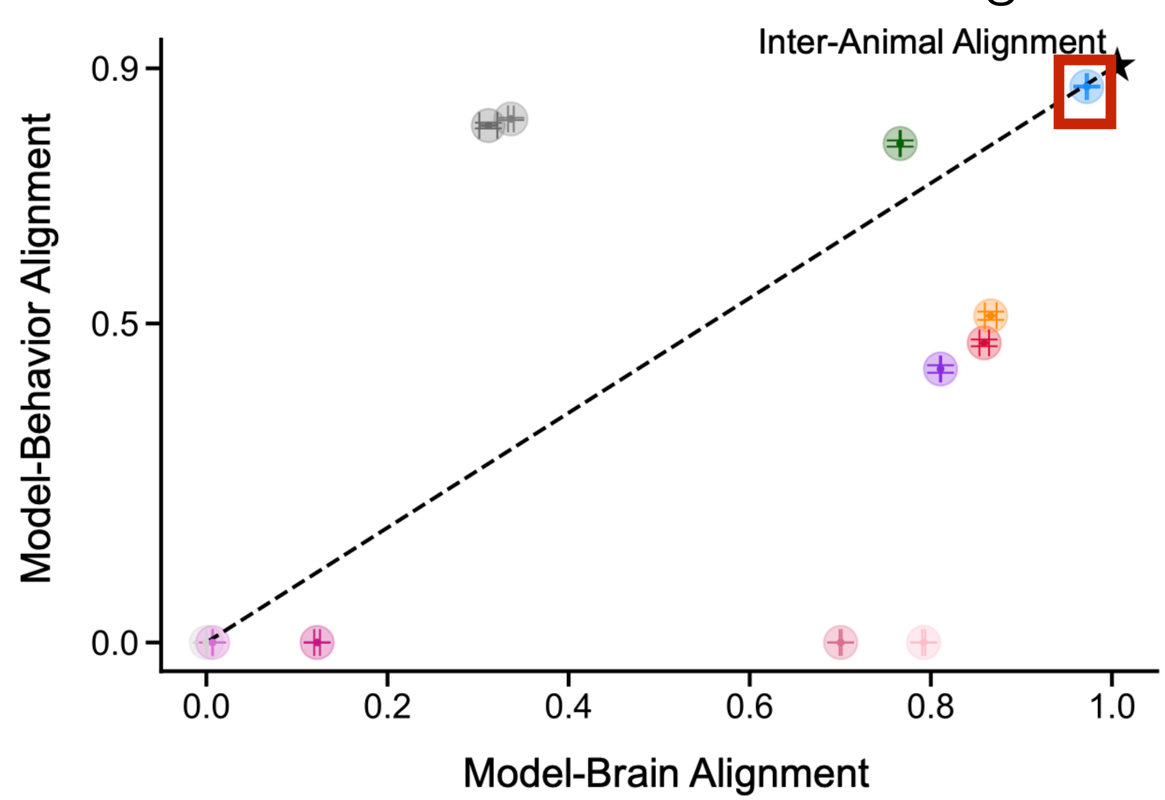
(and behavior)

Single-cell one-to-one alignment

Under a 1-1 mapping!

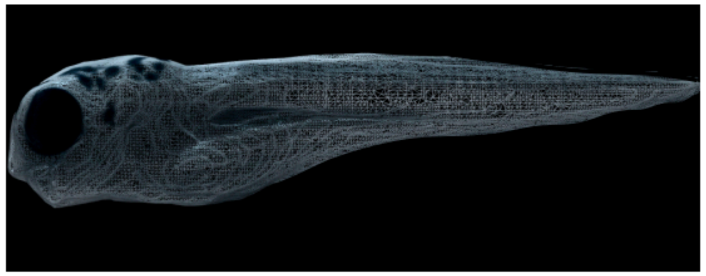


NeuroAI Turing Test

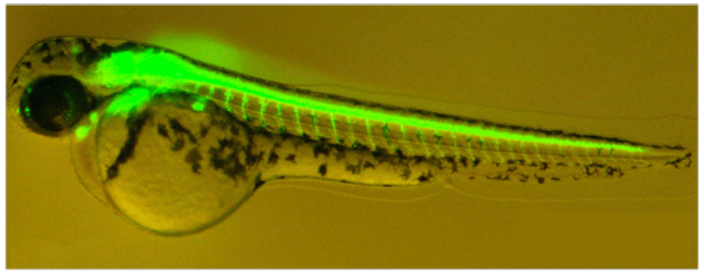


Recovers Mechanism

Agent



Zebrafish



Acknowledgements

NeuroAgents Lab

Contact:

anayebi@cs.cmu.edu

[@aran_nayebi](https://x.com/aran_nayebi)

https://cs.cmu.edu/~anayebi

Neuro Agents Lab

Carnegie Mellon
SCHOOL OF COMPUTER SCIENCE

Funding:

Foresight Institute

UK AISI Challenge Fund

Google Robotics Award

Burroughs Wellcome Fund CASI Award