Principled, Goal-Driven Models to Investigate Structure and Function in Neural Circuits

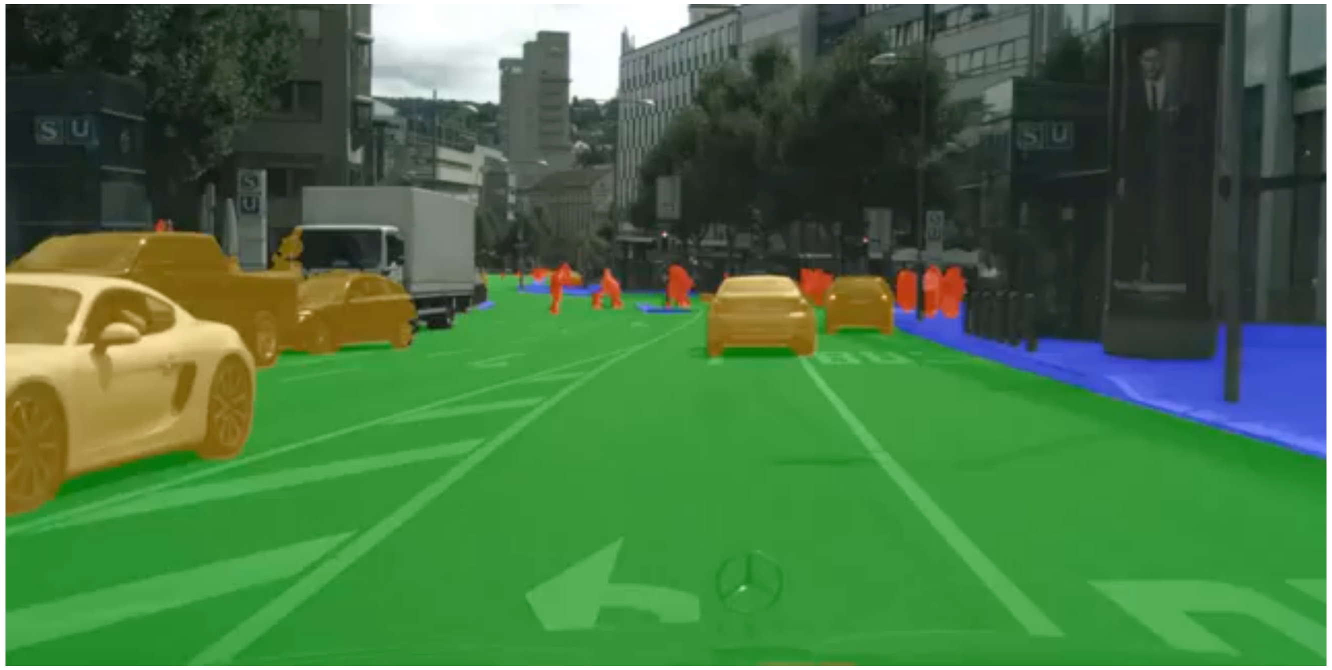
Aran Nayebi
McGovern Institute, MIT

Last updated: 2023.03.06

From Neurons to Behavior

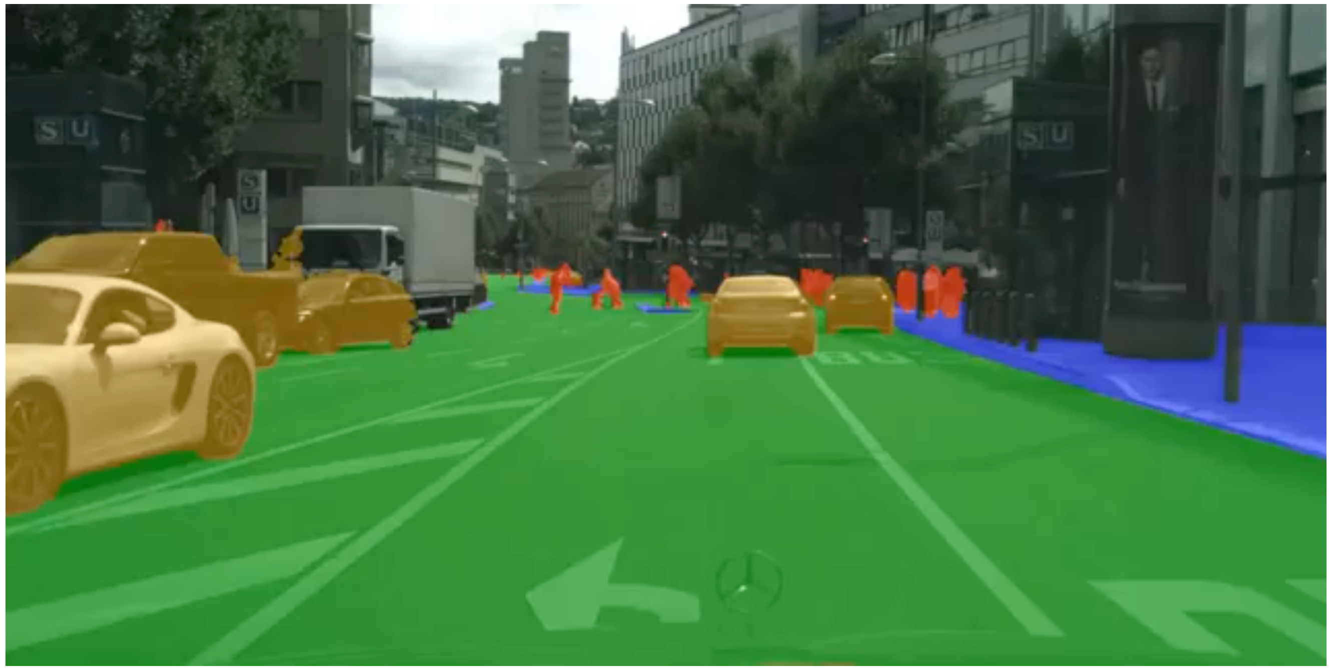
From Neurons to Behavior

Scene Understanding



From Neurons to Behavior

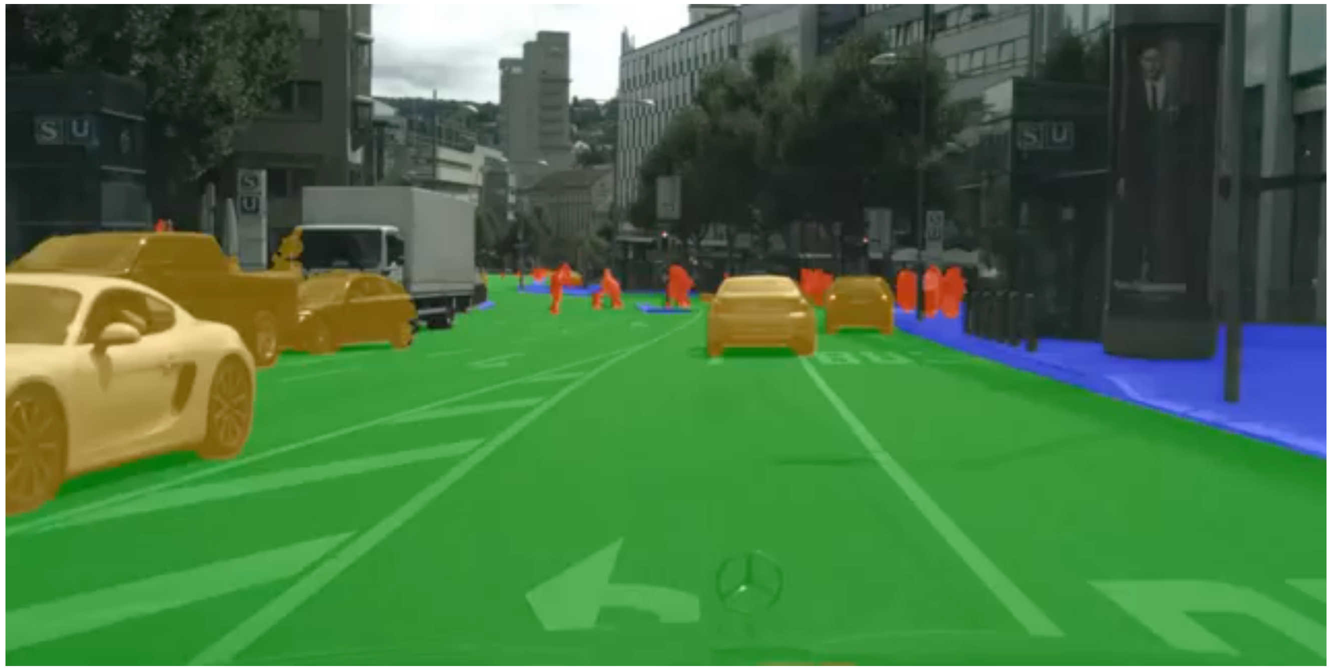
Scene Understanding



Multi-Step Planning

From Neurons to Behavior

Scene Understanding

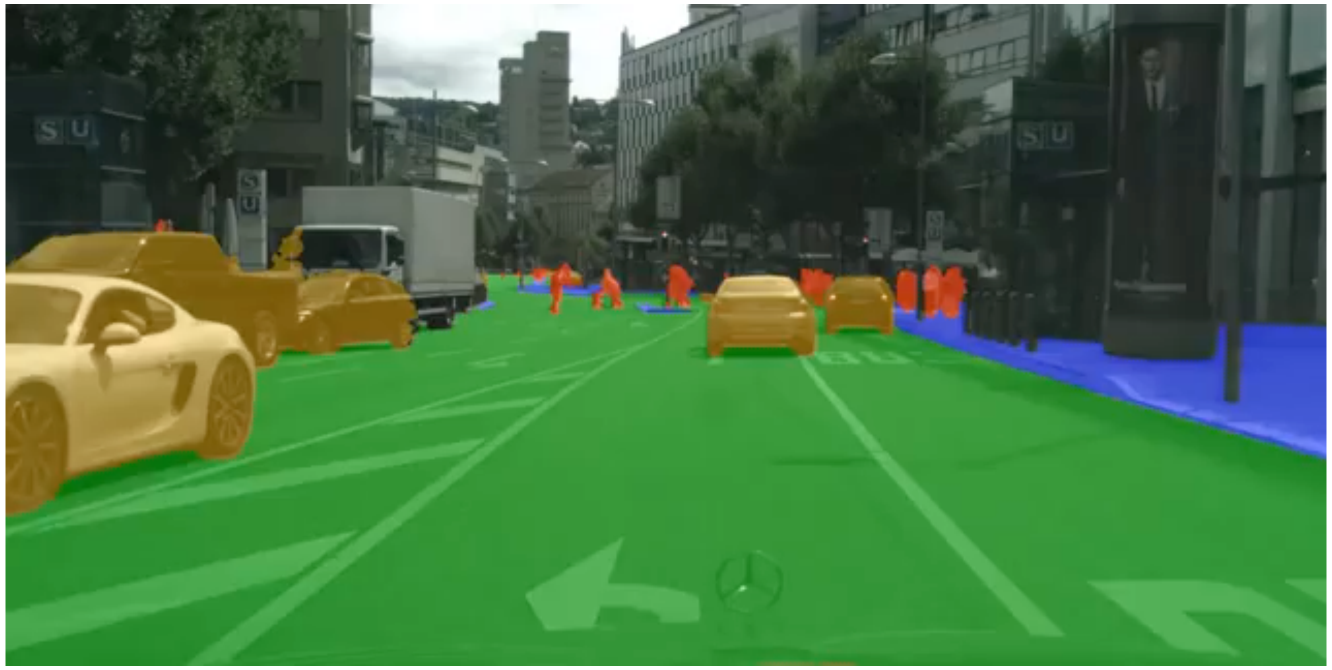


Multi-Step Planning

Navigation

From Neurons to Behavior

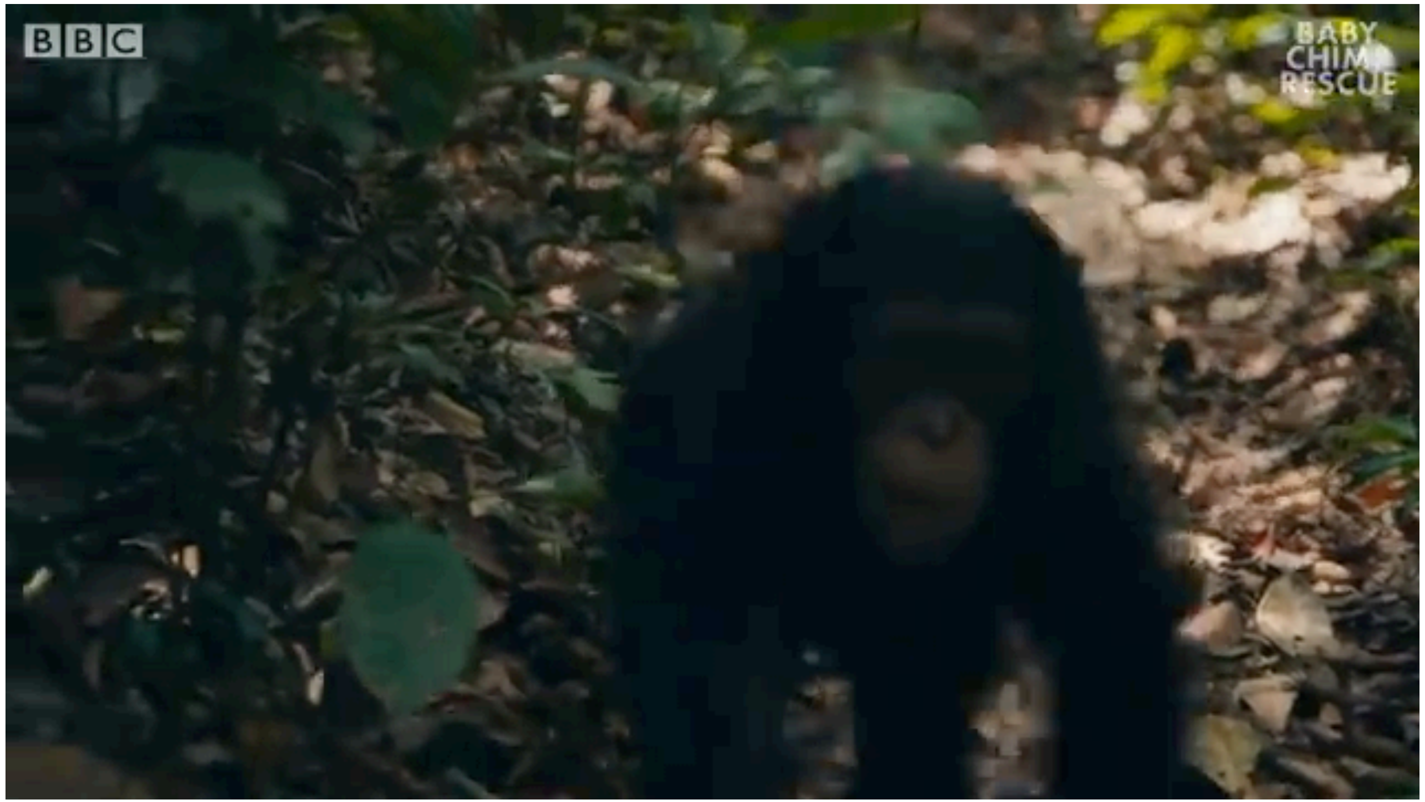
Scene Understanding



Multi-Step Planning

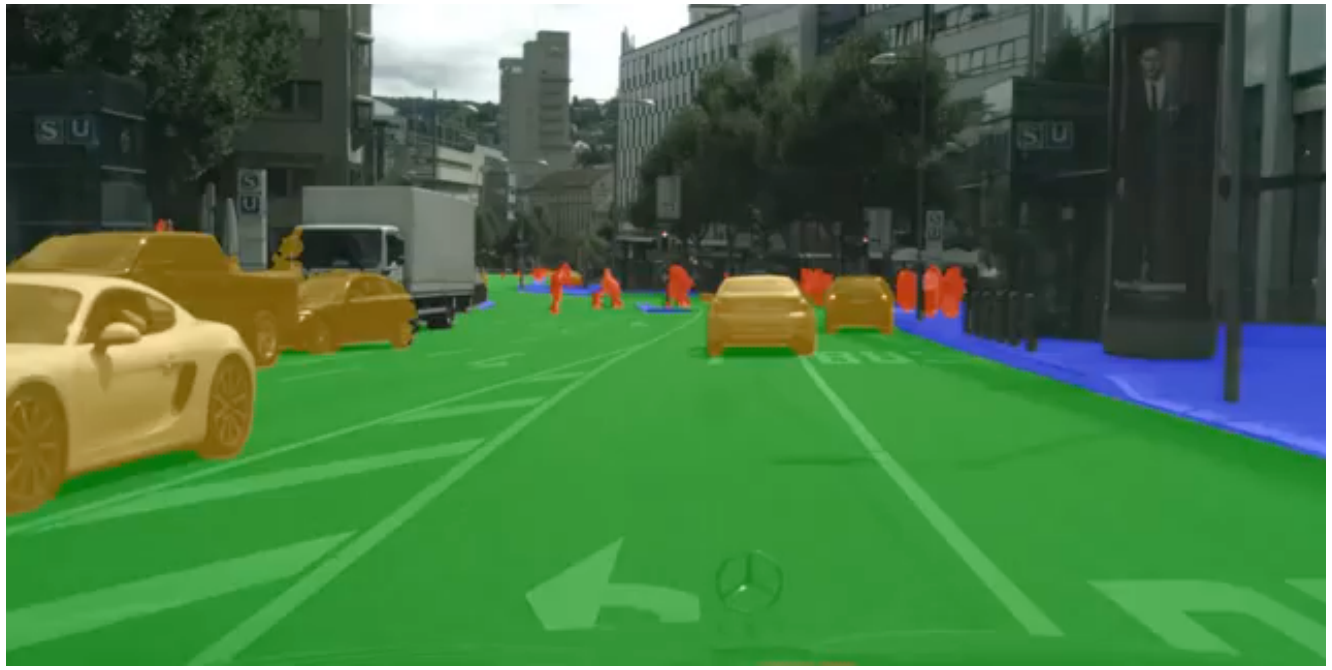
Navigation

Flexible Embodiment



From Neurons to Behavior

Scene Understanding

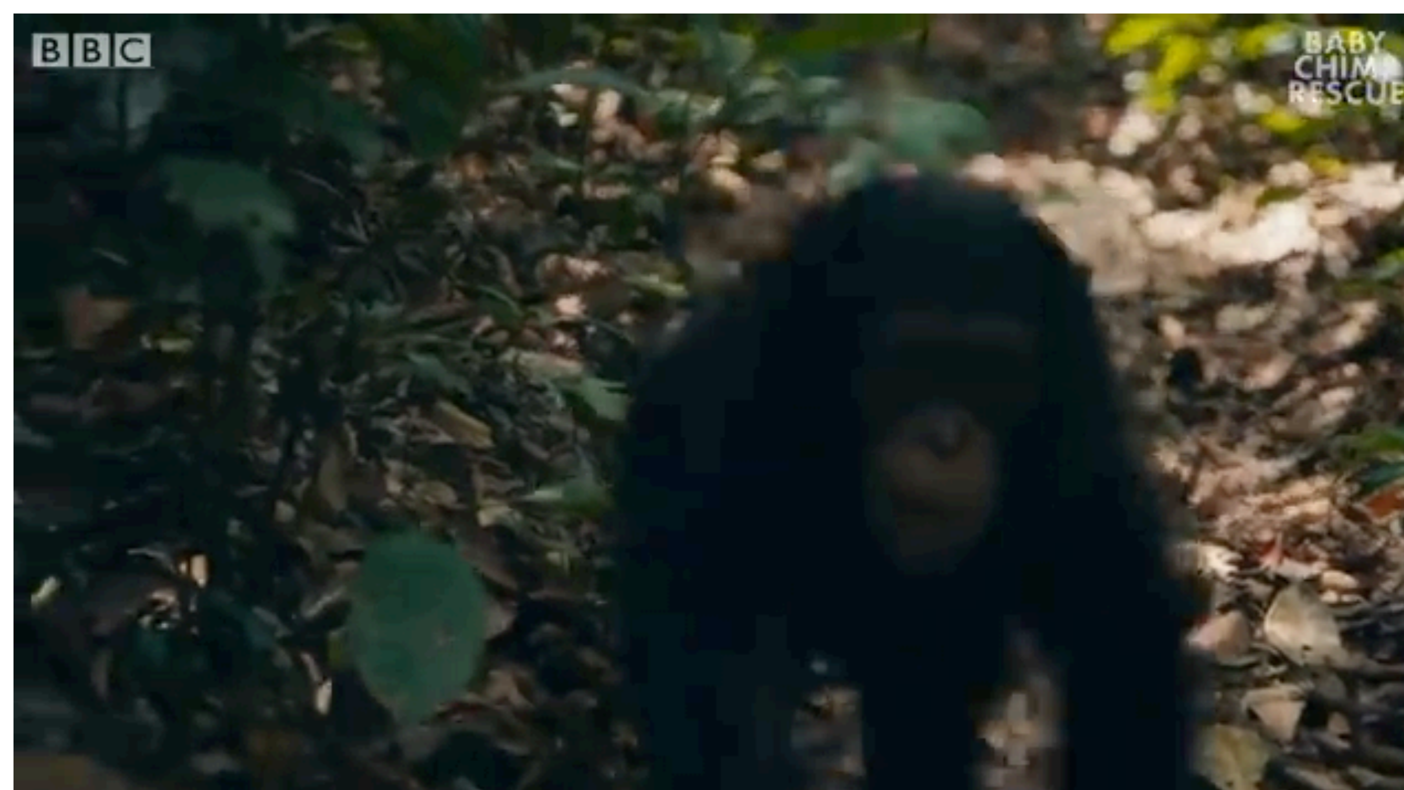


Multi-Step Planning

What are the core design principles that give rise to these abilities?

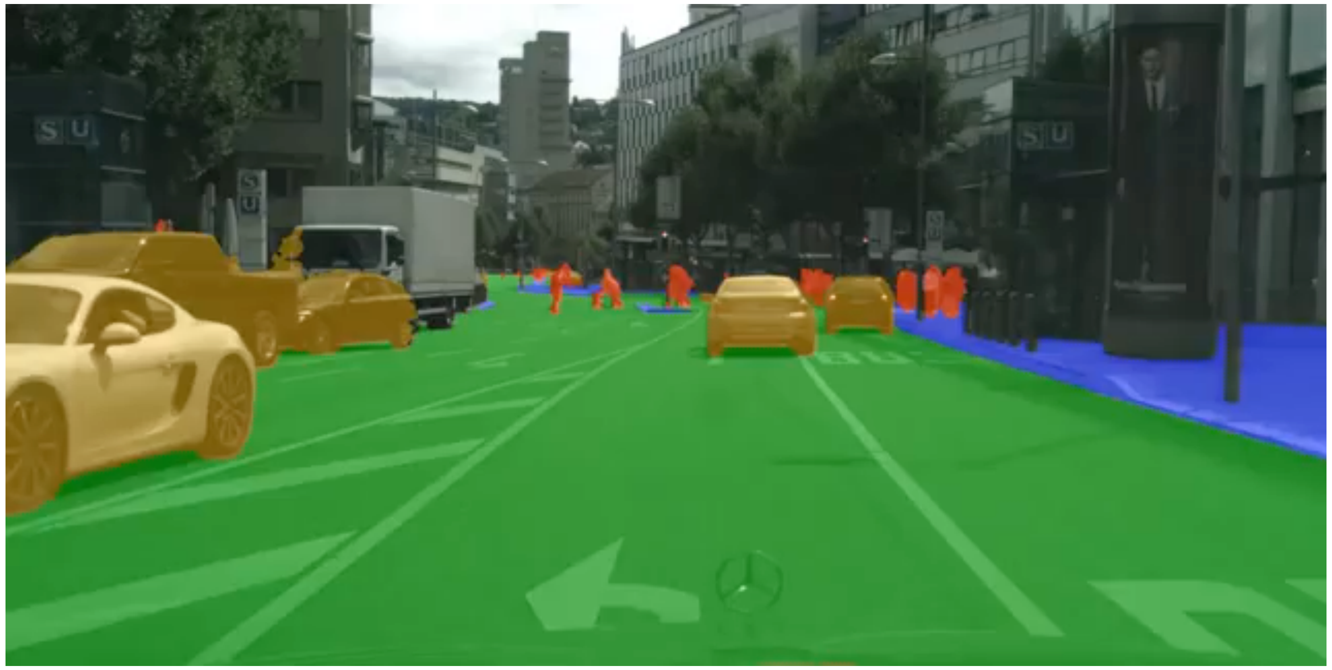
Navigation

Flexible Embodiment



From Neurons to Behavior

Scene Understanding

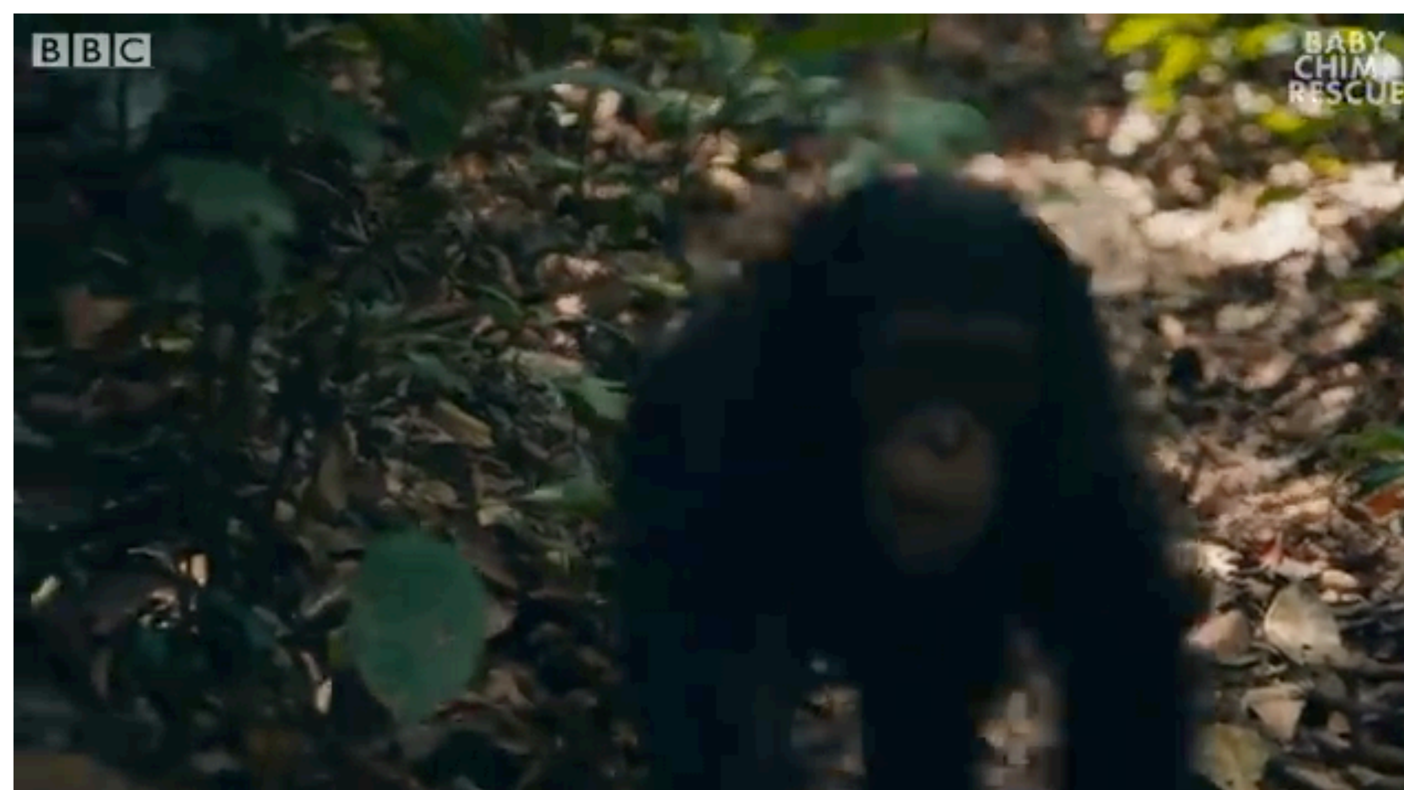


Multi-Step Planning

How do we bridge the gap from neurons to behavior?

Navigation

Flexible Embodiment

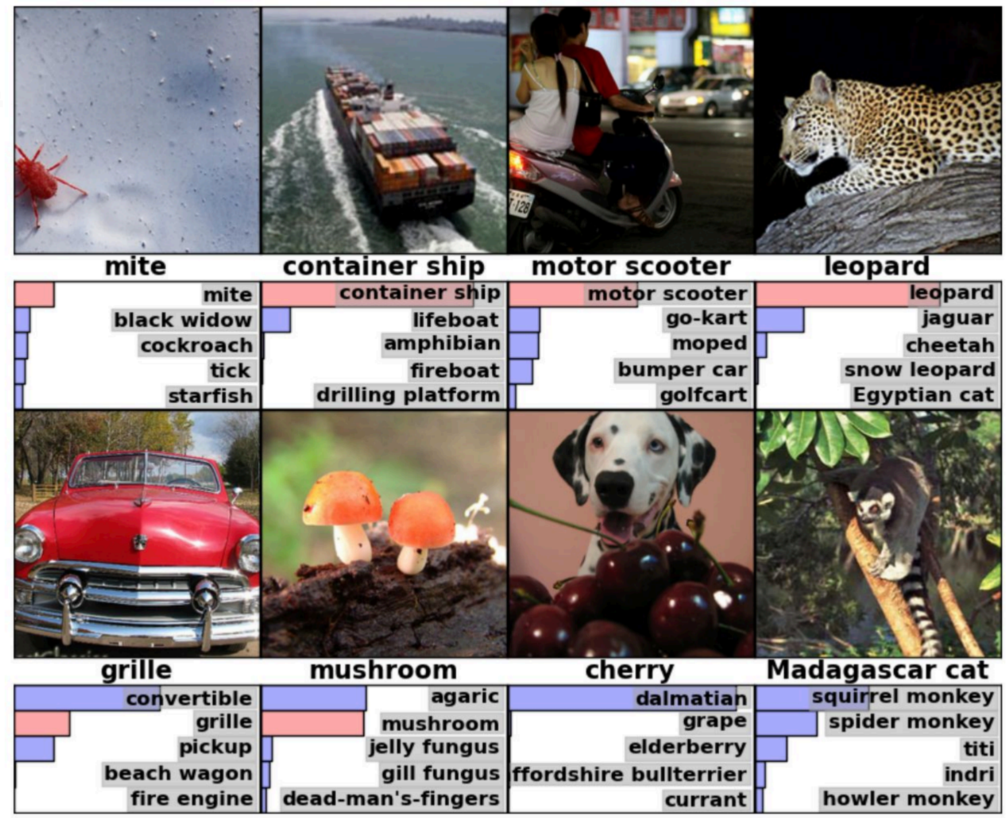


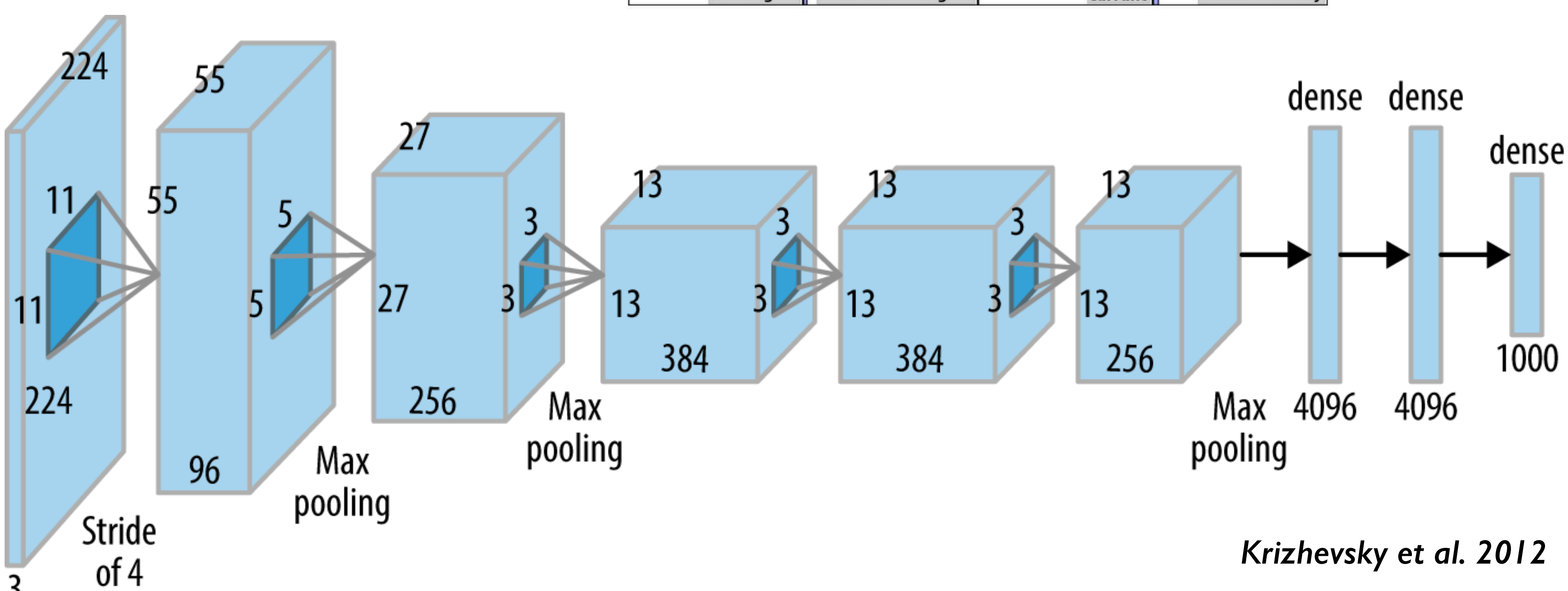
Functional Models: Convolutional Neural Networks (CNNs)

ImageNet Challenge

IMAGENET

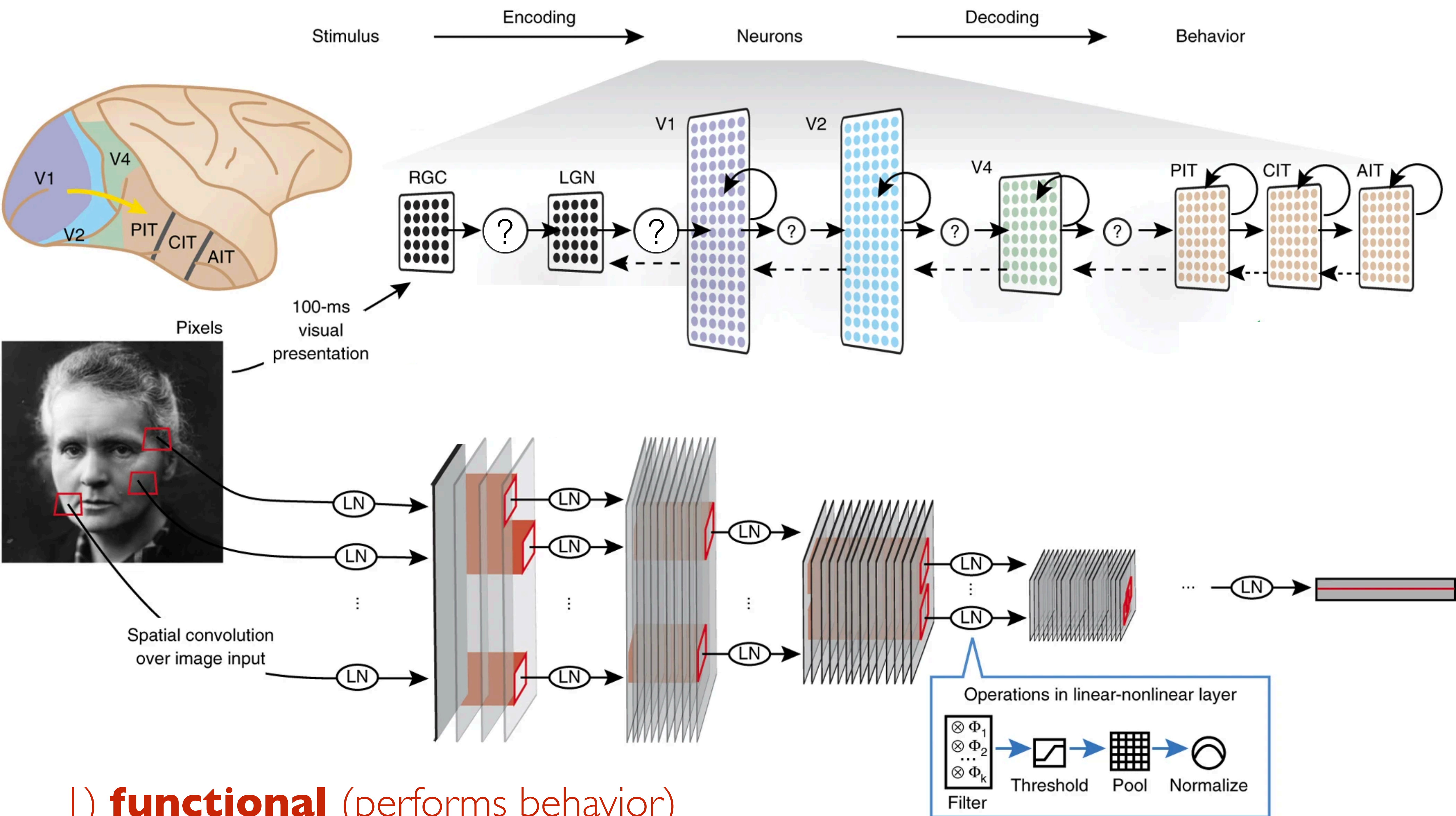
- 1,000 object classes (categories).
- Images:
  - 1.2 M train
  - 100k test.





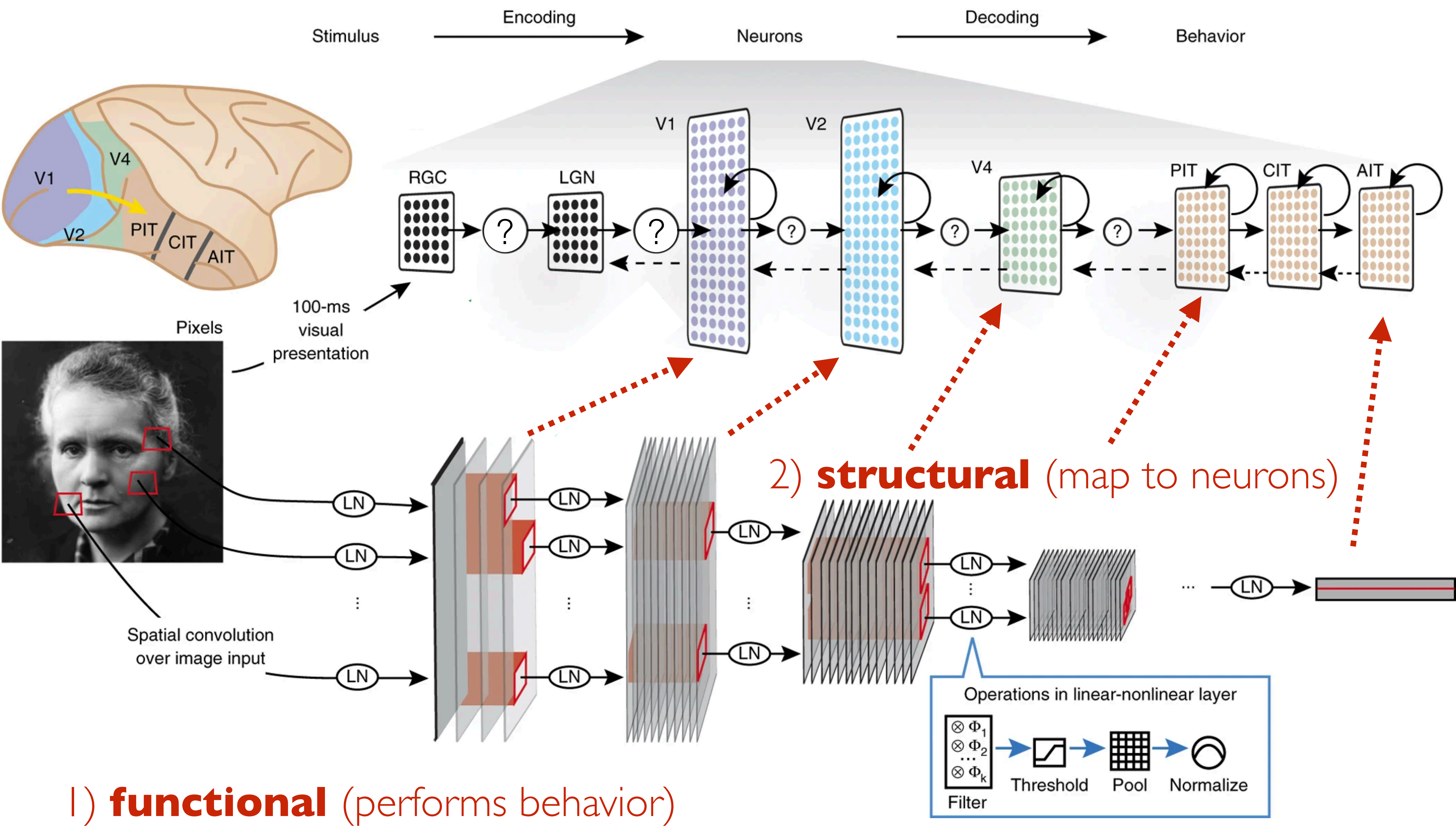
Krizhevsky et al. 2012

CNNs as Functional Models of Object Recognition



1) **functional** (performs behavior)

CNNs as *Structural* Models of Object Recognition



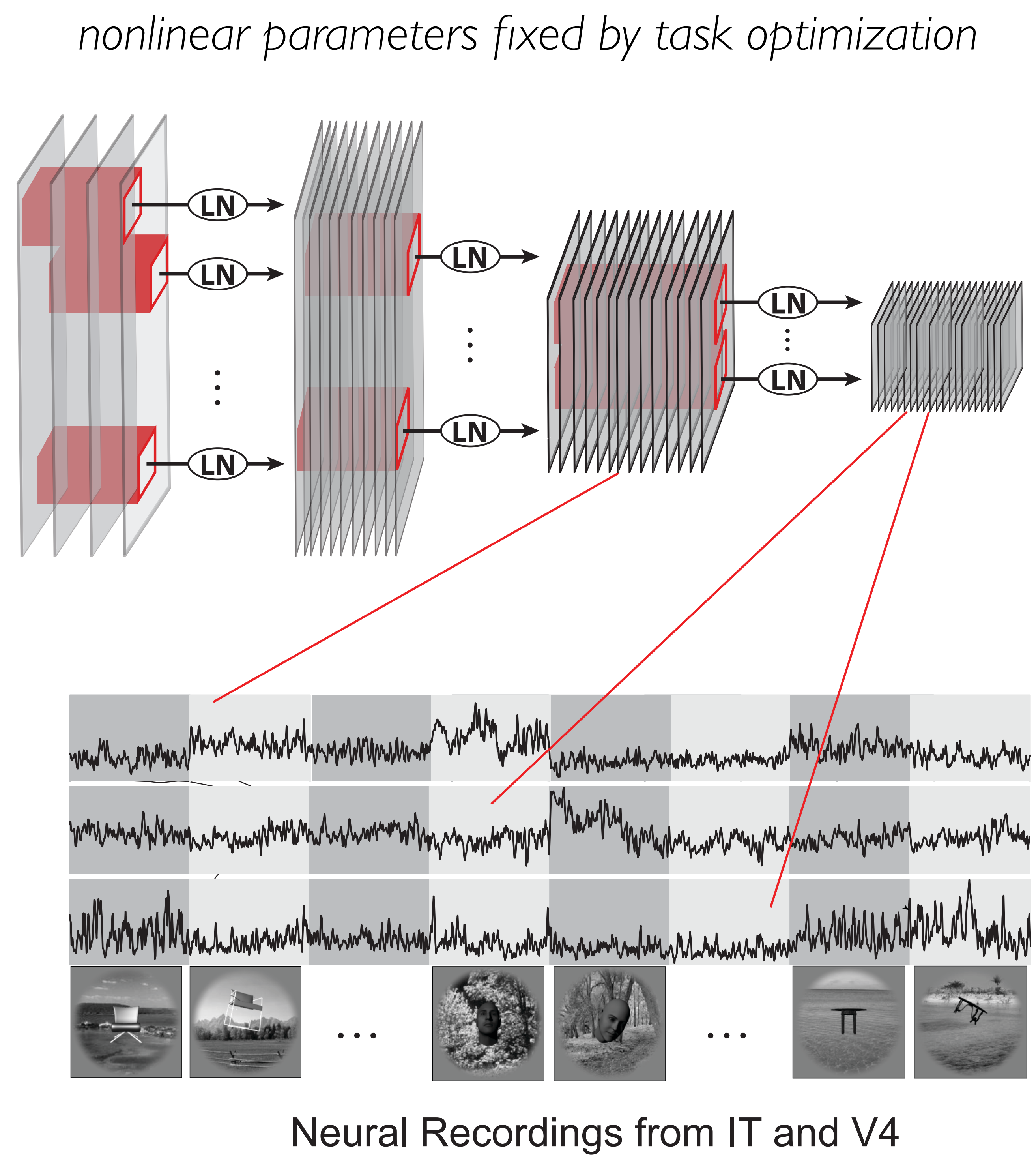
Neural predictivity: the ability of model to predict each individual neural site's activity.

Neural site unit ~ sparse linear combination of model units

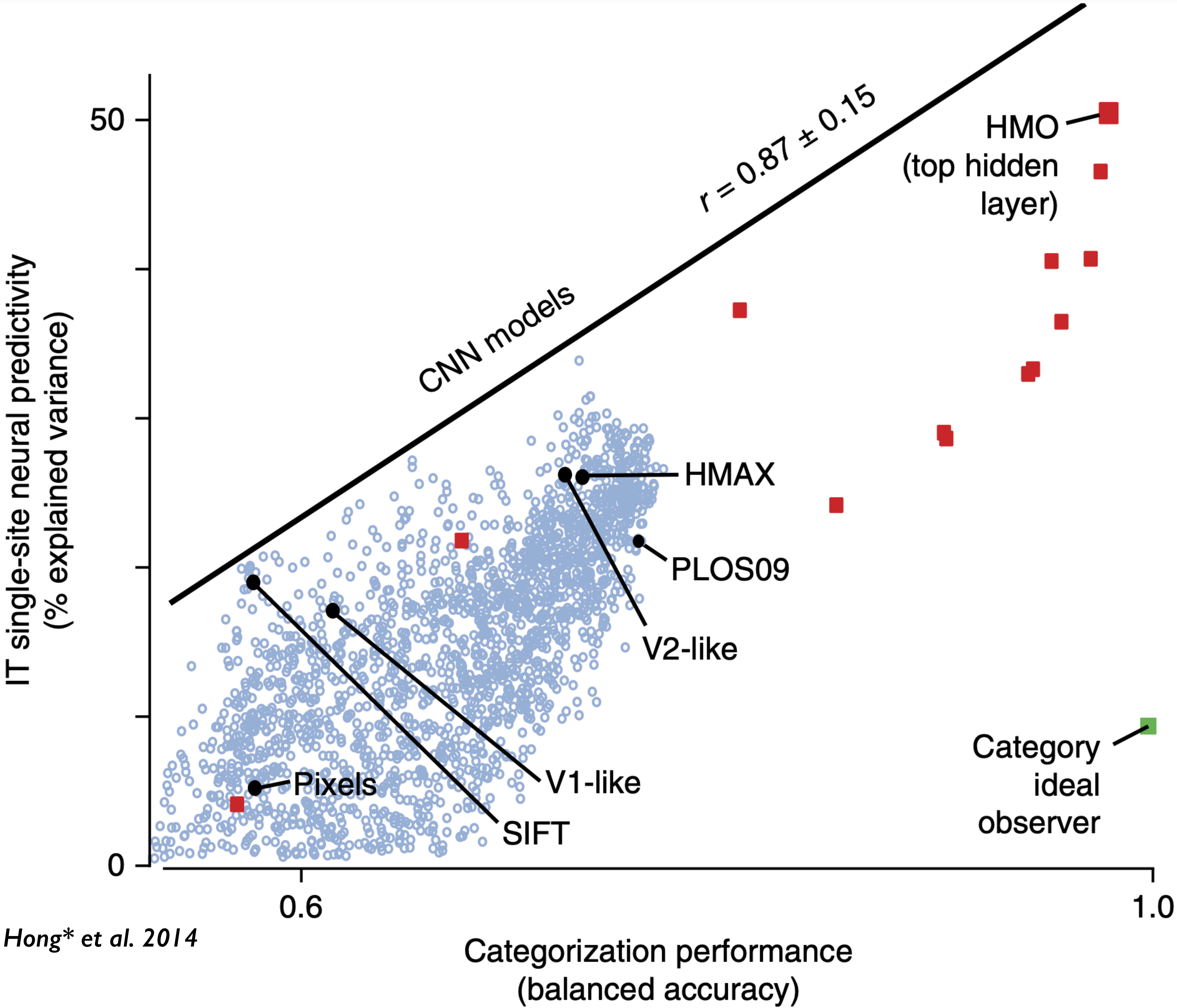
Linear regression with fixed training images.

Accuracy = goodness-of-fit on held-out testing images (Cross validated)

Neural predictivity = median accuracy over all units.

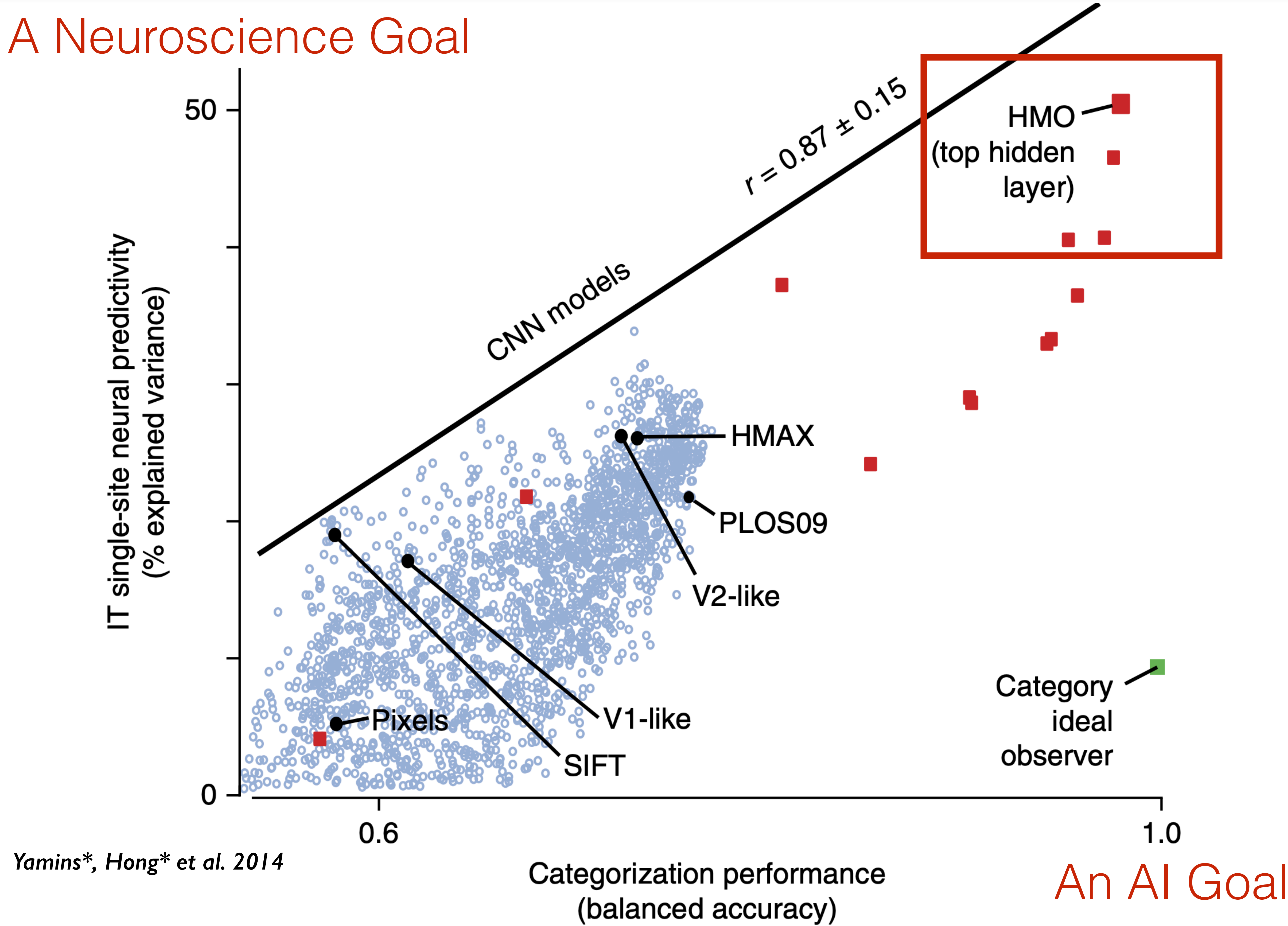


Task performance correlated with neural predictivity



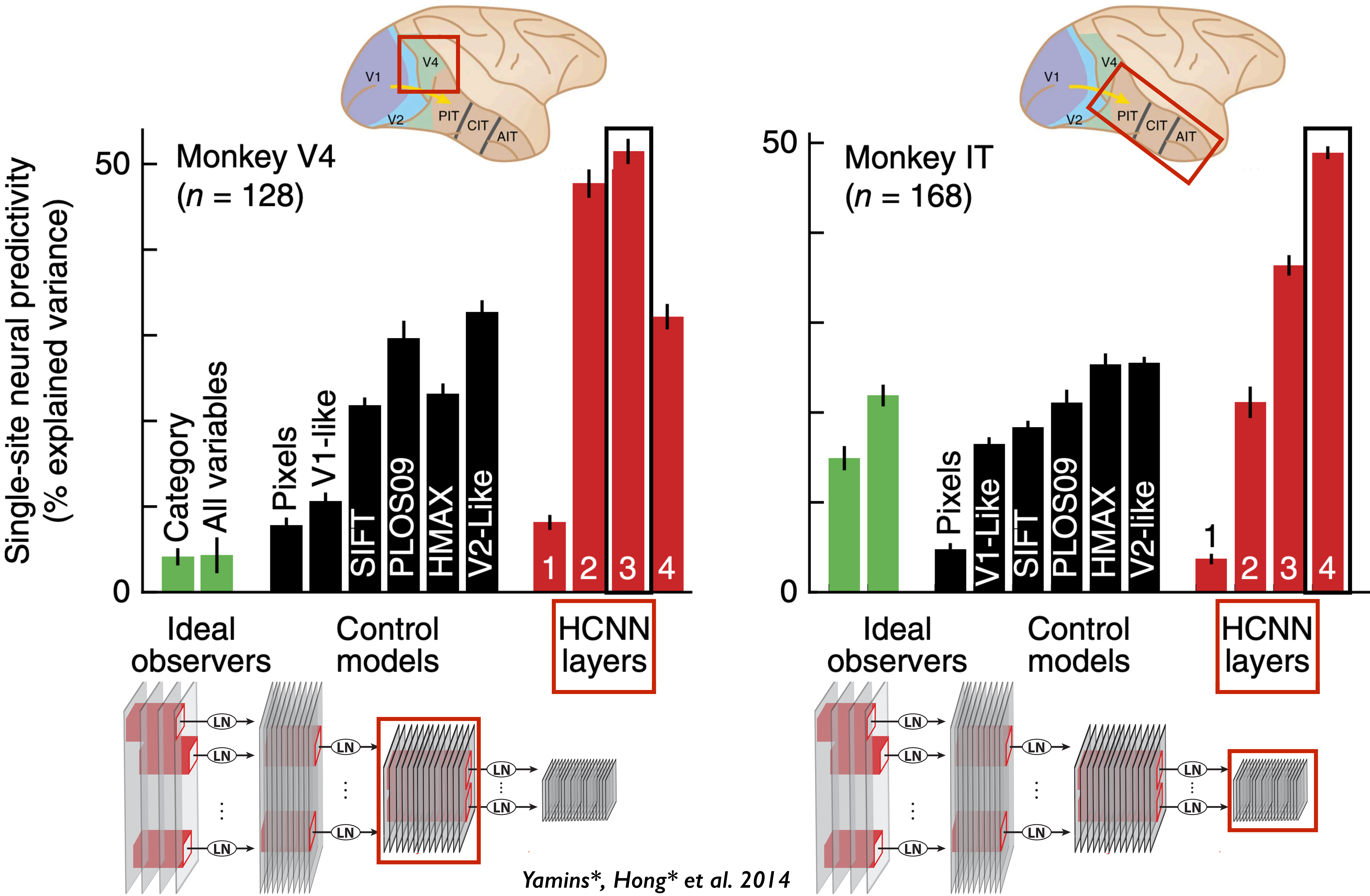
Yamins, Hong* et al. 2014*

Task performance correlated with neural predictivity



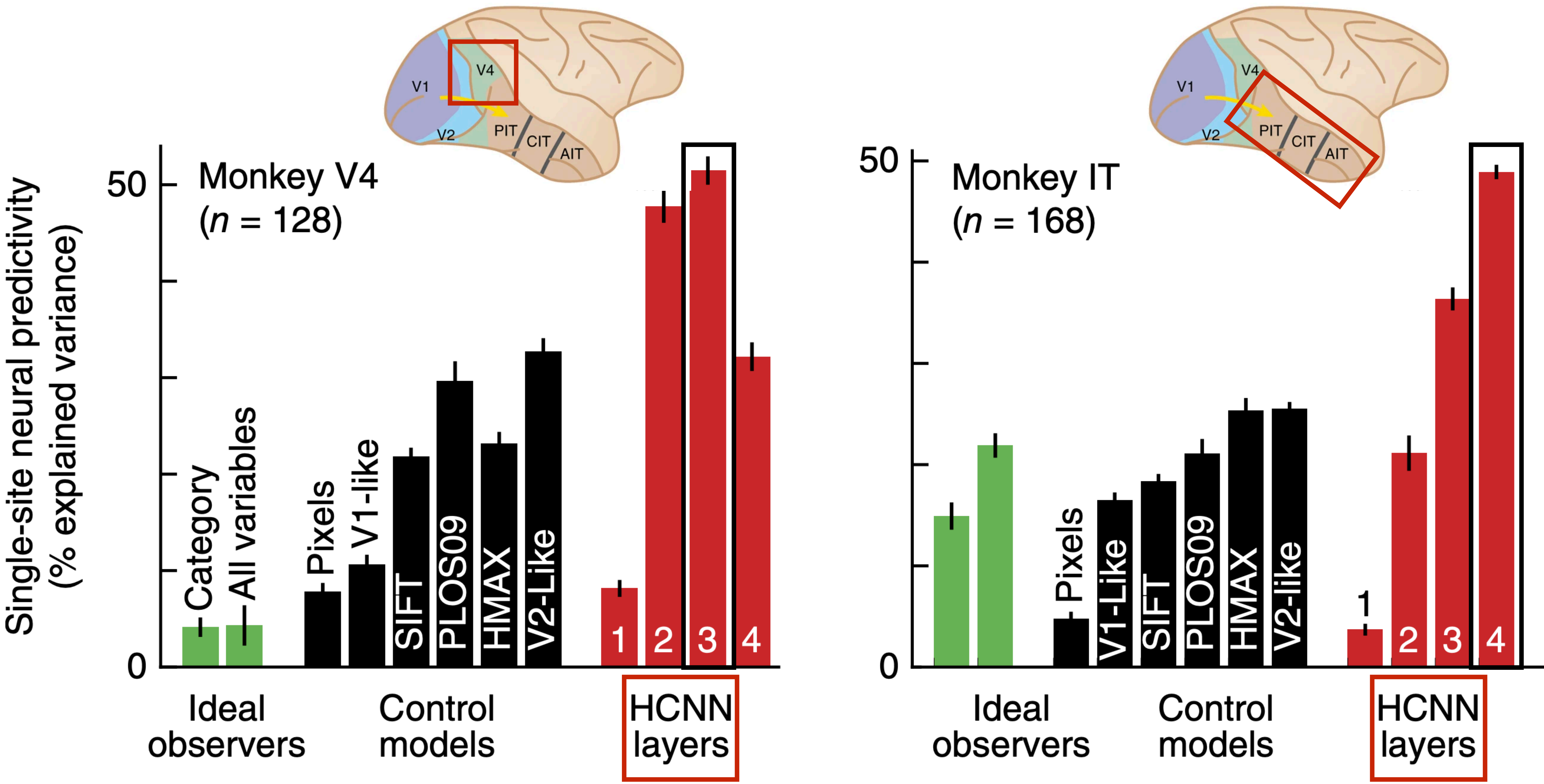
Yamins, Hong* et al. 2014*

Structural models as a by-product of task optimization



Yamins, Hong* et al. 2014*

Structural models as a by-product of task optimization



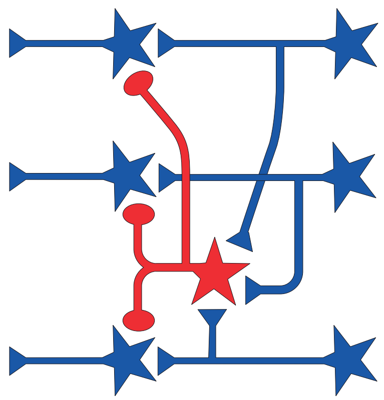
Is this an accident?

Goal-Driven Modeling - Three Design Principles

Goal-Driven Modeling - Three Design Principles

A = *architecture class*

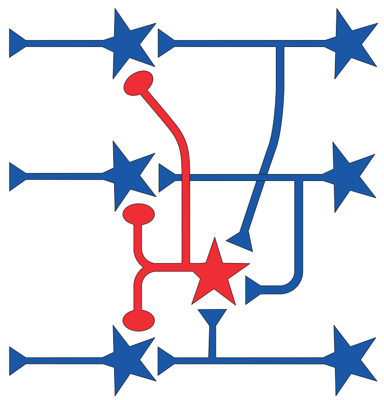
1. *"Circuit"*

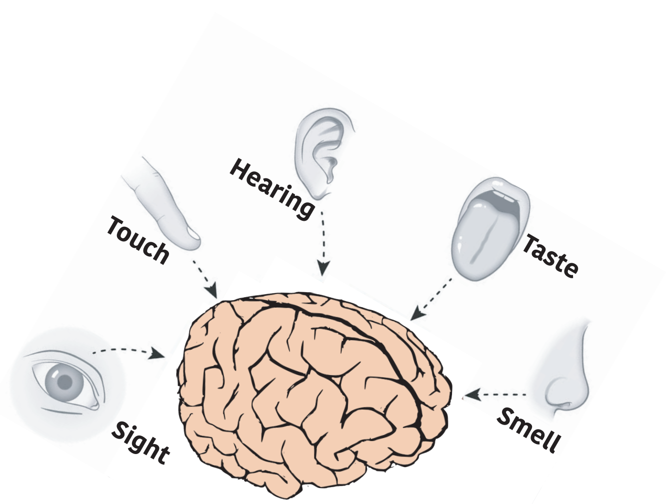


Goal-Driven Modeling - Three Design Principles

A = *architecture class*

1. *"Circuit"*





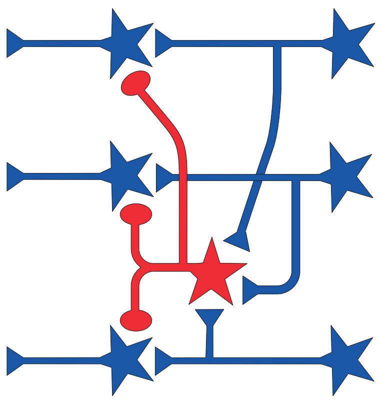
2. *"Environment"*

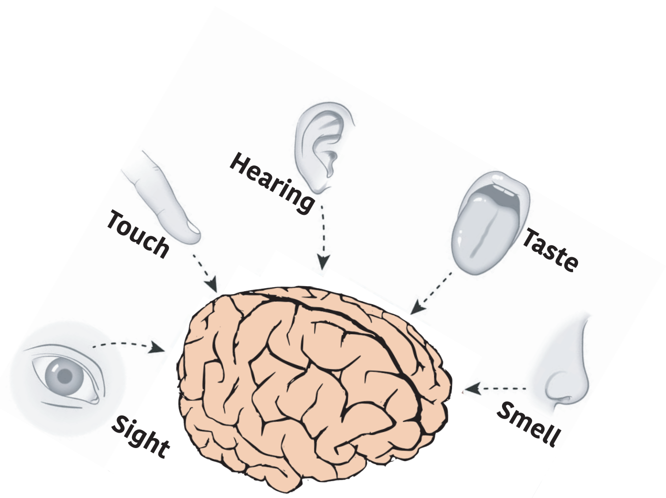
D = *data stream*

Goal-Driven Modeling - Three Design Principles

A = *architecture class*

1. *"Circuit"*





2. *"Environment"*

D = *data stream*

T = *task loss*

3. *"Ecological niche/behavior"*

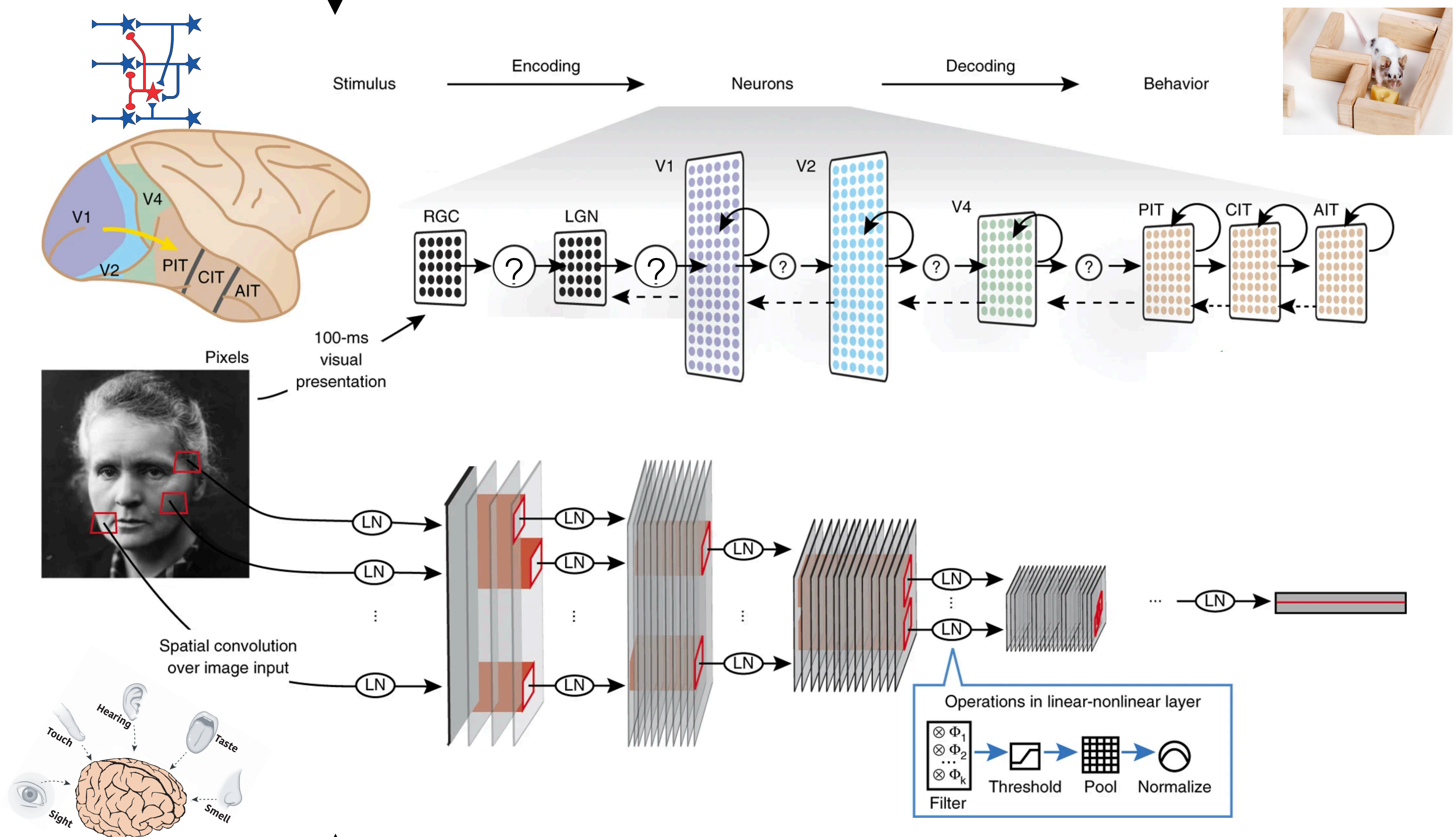
Goal-Driven Modeling - Three Design Principles

A = *architecture class*

1. *"Circuit"*

T = *task loss*

3. *"Ecological niche/behavior"*



2. *"Environment"*

D = *data stream*

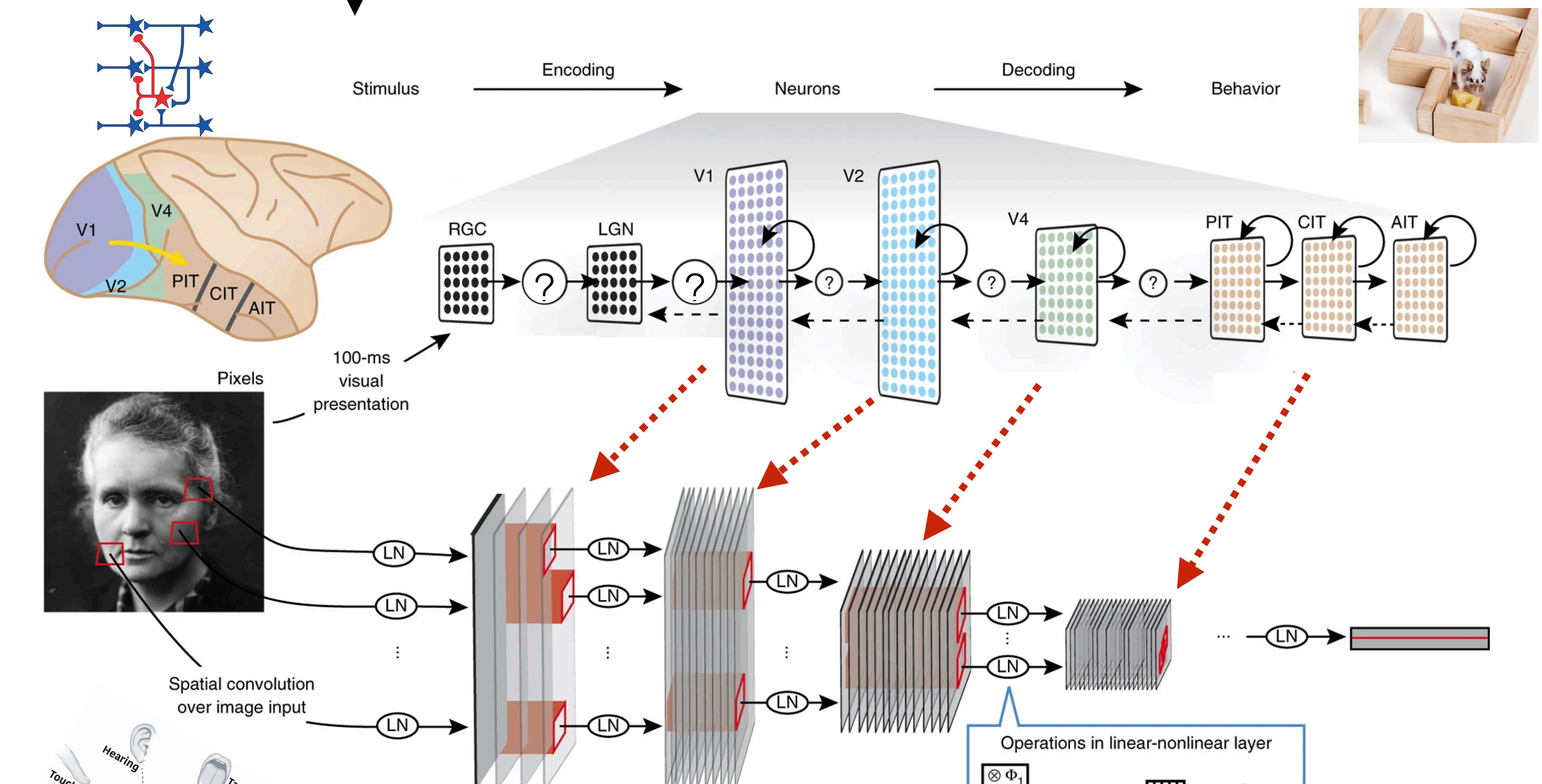
Goal-Driven Modeling - Three Design Principles

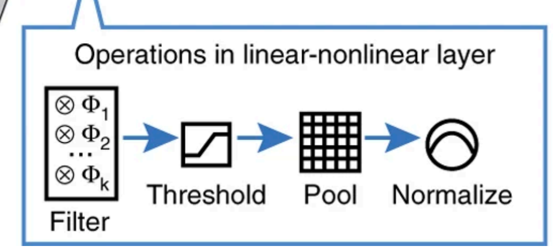
A = *architecture class*

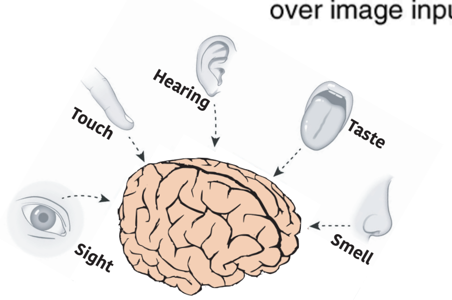
1. *"Circuit"*

T = *task loss*

3. *"Ecological niche/behavior"*







2. *"Environment"*

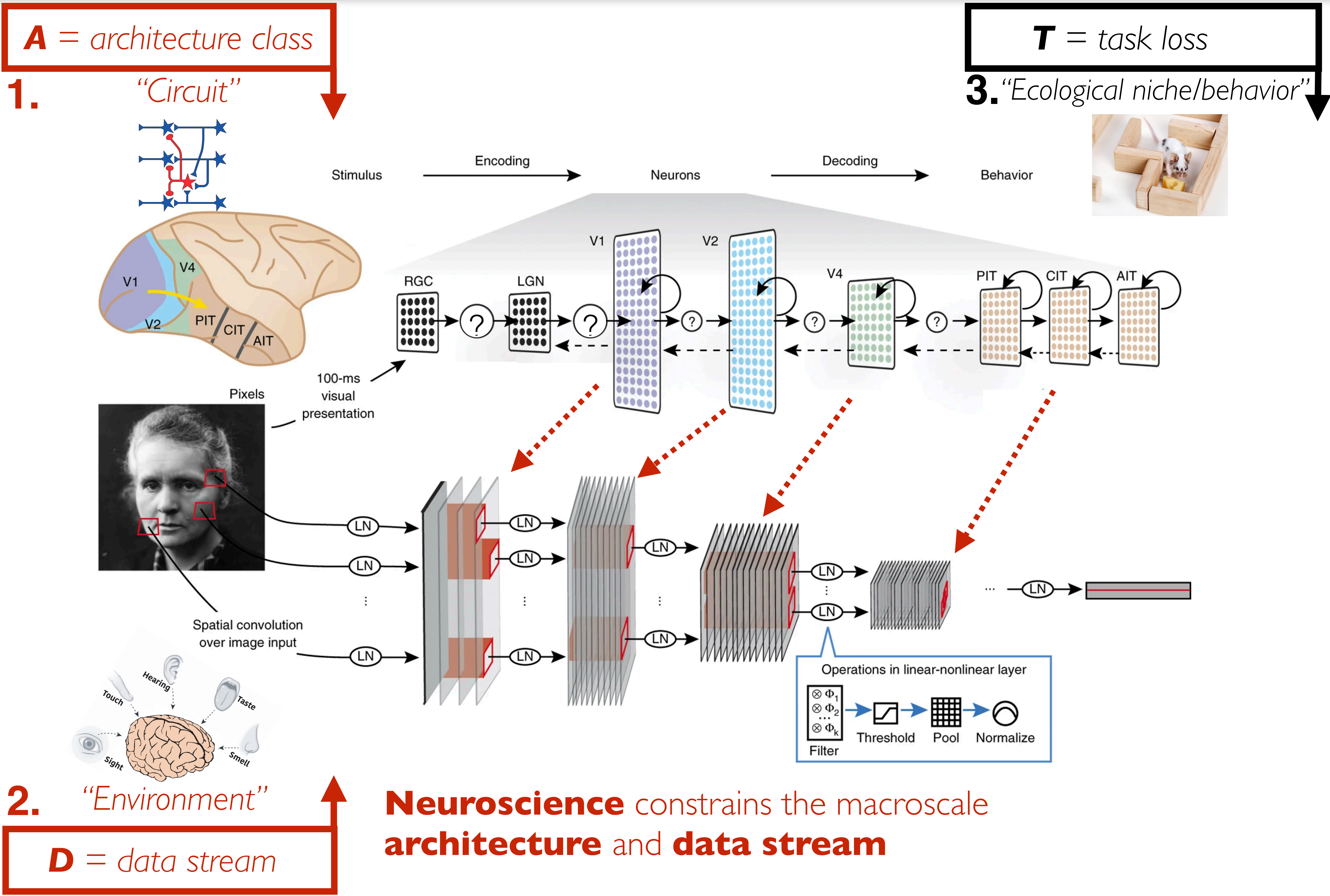
D = *data stream*

CNNs are inspired by visual neuroscience:

1) **hierarchy**

2) **retinotopy** (spatially tiled)

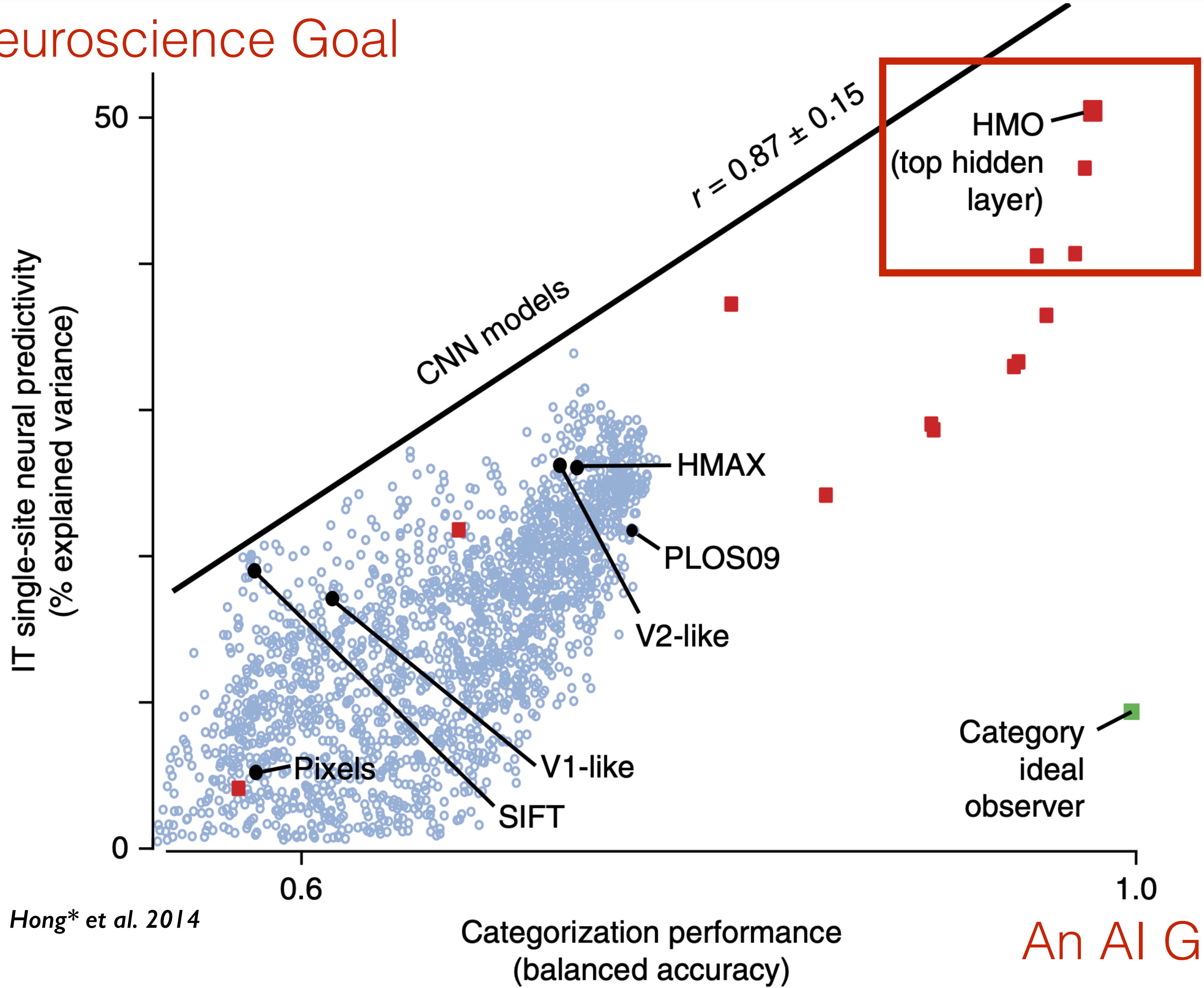
Goal-Driven Modeling - Three Design Principles



Neuroscience constrains the macroscale **architecture** and **data stream**

Task performance correlated with neural predictivity

A Neuroscience Goal

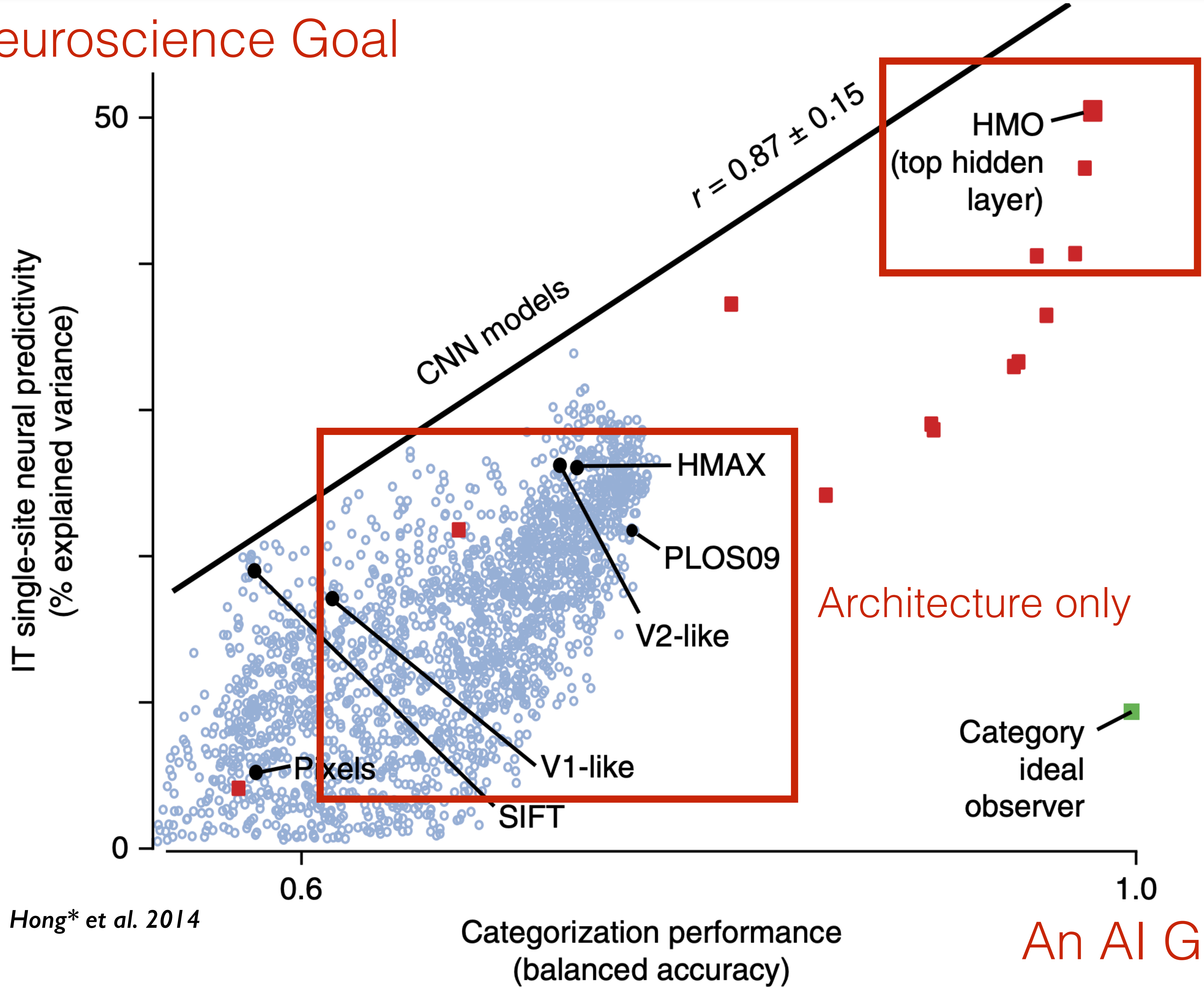


Yamins, Hong* et al. 2014*

An AI Goal

Task performance correlated with neural predictivity

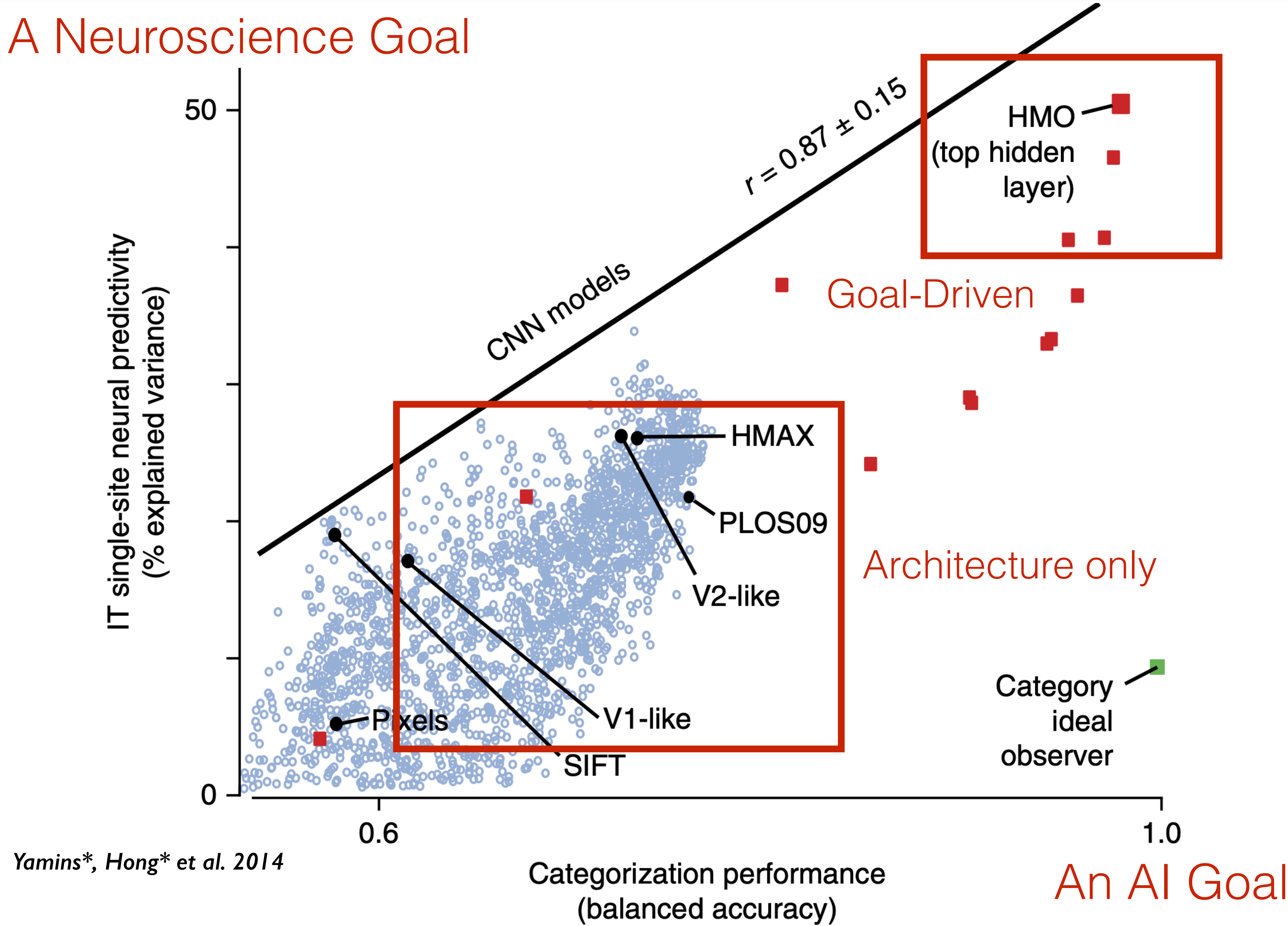
A Neuroscience Goal



Yamins, Hong* et al. 2014*

An AI Goal

Task performance correlated with neural predictivity



Yamins, Hong* et al. 2014*

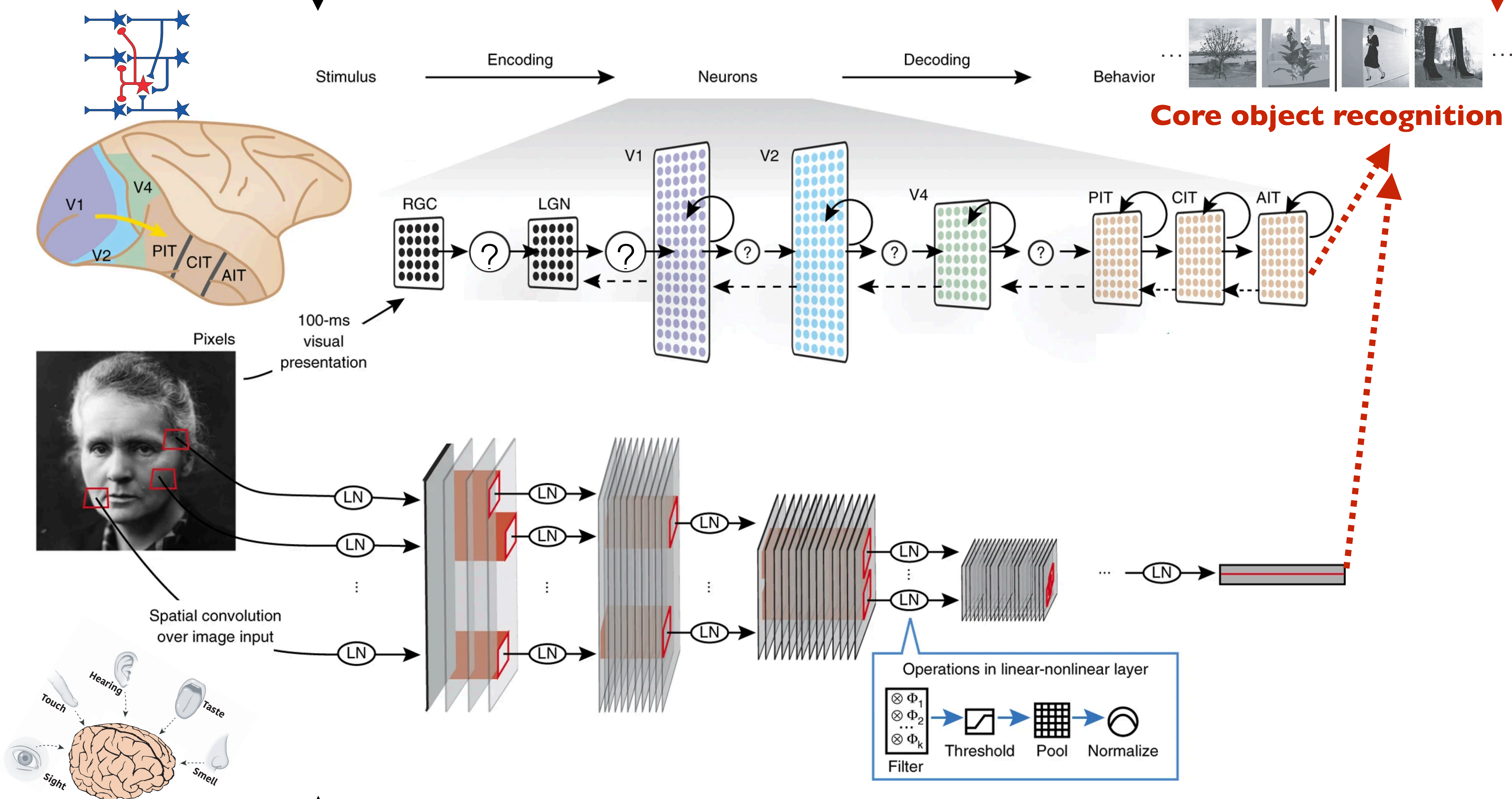
Goal-Driven Modeling - Three Design Principles

A = *architecture class*

1. *"Circuit"*

T = *task loss*

3. *"Ecological niche/behavior"*



2. *"Environment"*

D = *data stream*

Cognitive science sets the **task loss**

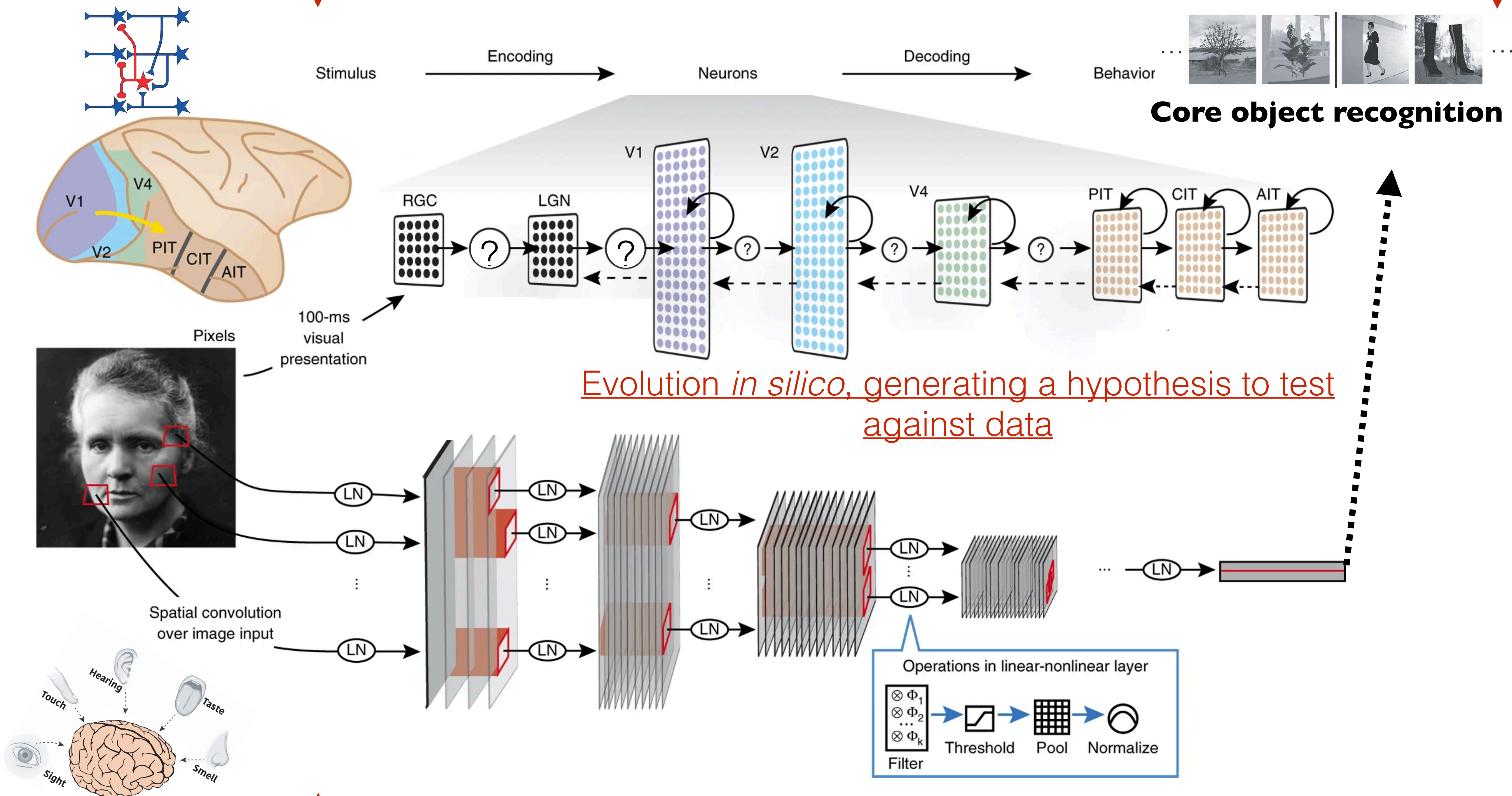
Goal-Driven Modeling - Three Design Principles

A = *architecture class*

1. *"Circuit"*

T = *task loss*

3. *"Ecological niche/behavior"*



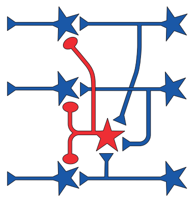
2. *"Environment"*

D = *data stream*

Reverse-Engineering Natural Intelligence

A = *architecture class*

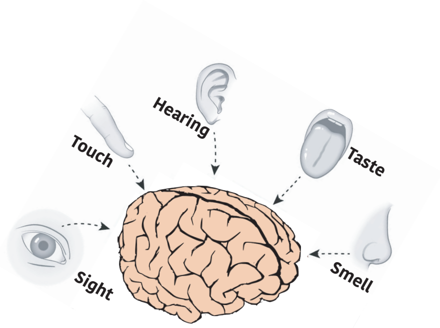
1. *"Circuit"*



T = *task loss*

3. *"Ecological niche/behavior"*

I will show that this *general* approach yields a normative understanding of evolutionary constraints *across* species (rodents & primates), and sensory and *non-sensory* areas



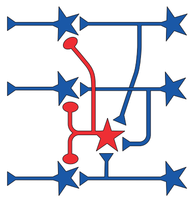
2. *"Environment"*

D = *data stream*

Reverse-Engineering Natural Intelligence

A *= architecture class*

1. *"Circuit"*



Neuroscience constrained

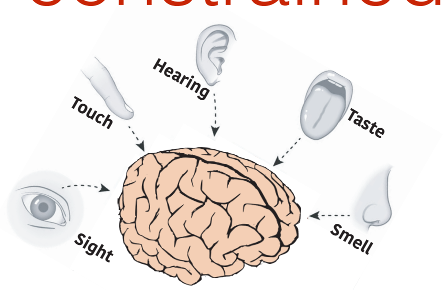
T *= task loss*

3. *"Ecological niche/behavior"*

Cognitive science constrained

I will show that this *general* approach yields a normative understanding of evolutionary constraints *across* species (rodents & primates), and sensory and *non-sensory* areas

Neuroscience constrained



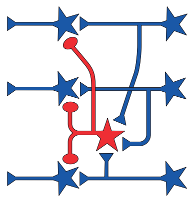
2. *"Environment"*

D *= data stream*

Reverse-Engineering Natural Intelligence

A = *architecture class*

1. *"Circuit"*



Neuroscience constrained

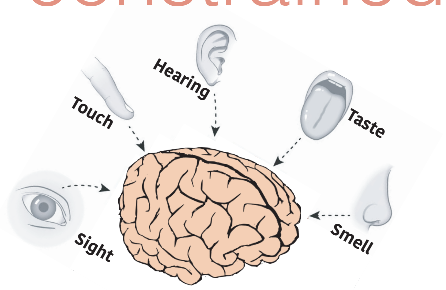
T = *task loss*

3. *"Ecological niche/behavior"*

Cognitive science constrained

I will show that this *general* approach yields a normative understanding of evolutionary constraints *across* species (rodents & primates), and sensory and *non-sensory* areas

Neuroscience constrained



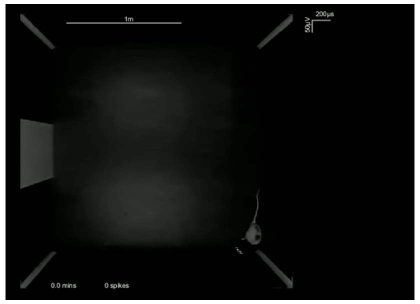
2. *"Environment"*

D = *data stream*

- Neurobiological Puzzle
- Core Conceptual Insights

Reverse-Engineering Strategy: Bridging Neurons to Behavior

Reverse-Engineering Strategy: Bridging Neurons to Behavior



Navigation

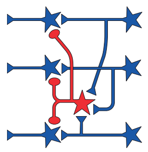
1. <u>Operationalize</u> a behavioral domain of interest.
Ways in which brains surpass current engineered systems? (e.g. performance, size, energy, etc)

Core object recognition

Reverse-Engineering Strategy: Bridging Neurons to Behavior

A = *architecture class*

1. *"Circuit"*

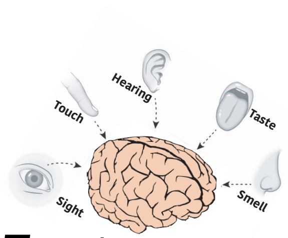


T = *task loss*

3. *"Ecological niche/behavior"*

1. Operationalize a behavioral domain of interest.
Ways in which brains surpass current engineered systems? (e.g. performance, size, energy, etc)

2. Hypothesize architectures and tasks (loss functions).
Optimize for task (evolution *in silico*).
Predict held out neural & behavioral data.



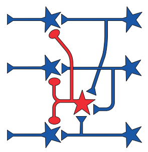
2. *"Environment"*

D = *data stream*

Reverse-Engineering Strategy: Bridging Neurons to Behavior

A = *architecture class*

1. *"Circuit"*



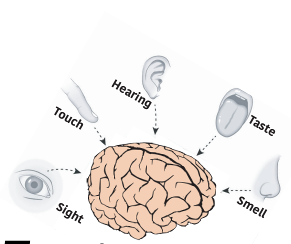
T = *task loss*

3. *"Ecological niche/behavior"*

1. <u>Operationalize</u> a behavioral domain of interest.
Ways in which brains surpass current engineered systems? (e.g. performance, size, energy, etc)

2. <u>Hypothesize</u> architectures and tasks (loss functions).
<u>Optimize</u> for task (evolution *in silico*).
<u>*Predict* held out</u> neural & behavioral data.

<u>3. Core Conceptual Insights:</u>
<u>Identify</u> the patterns of the best architectures & tasks.



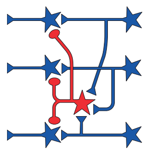
2. *"Environment"*

D = *data stream*

Reverse-Engineering Strategy: Bridging Neurons to Behavior

A = *architecture class*

1. *"Circuit"*



T = *task loss*

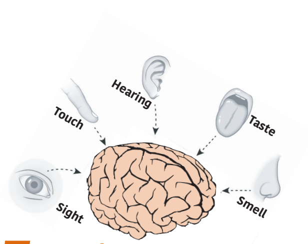
3. *"Ecological niche/behavior"*

1. Operationalize a behavioral domain of interest. Ways in which brains surpass current engineered systems? (e.g. performance, size, energy, etc)

2. Hypothesize architectures and tasks (loss functions). Optimize for task (evolution *in silico*). *Predict* held out neural & behavioral data.

3. Core Conceptual Insights: Identify the patterns of the best architectures & tasks.

A structural & functional *normative* understanding of the evolutionary constraints of the biological system to produce the behavior in (1).



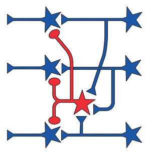
2. *"Environment"*

D = *data stream*

Reverse-Engineering Strategy: Bridging Neurons to Behavior

A *= architecture class*

1. *"Circuit"*



3. *"Ecological niche/behavior"*

T *= task loss*

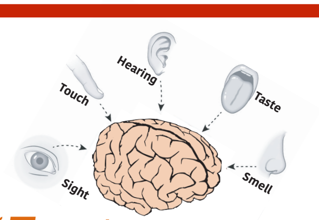
1. Operationalize a behavioral domain of interest. Ways in which brains surpass current engineered systems? (e.g. performance, size, energy, etc)

2. Hypothesize architectures and tasks (loss functions). Optimize for task (evolution *in silico*). *Predict* held out neural & behavioral data.

3. Core Conceptual Insights:

Identify the patterns of the best architectures & tasks.

A structural & functional *normative* understanding of the evolutionary constraints of the biological system to produce the behavior in (1).



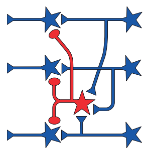
2. *"Environment"*

D *= data stream*

Reverse-Engineering Strategy: Bridging Neurons to Behavior

A = *architecture class*

1. "Circuit"



T = *task loss*

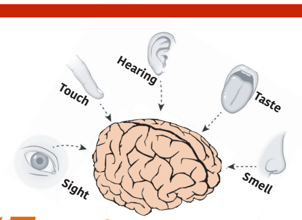
3. "Ecological niche/behavior"

1. Operationalize a behavioral domain of interest. Ways in which brains surpass current engineered systems? (e.g. performance, size, energy, etc)

2. Hypothesize architectures and tasks (loss functions). Optimize for task (evolution *in silico*). *Predict held out* neural & behavioral data.

3. Core Conceptual Insights:
Identify the patterns of the best architectures & tasks.

A structural & functional *normative* understanding of the evolutionary constraints of the biological system to produce the behavior in (1).



2. "Environment"

D = *data stream*

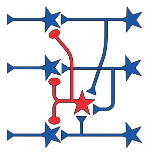
Applications:

Predictions of causal perturbations, interventions
Perceptual control/prediction (BMI)
Hypothesis generation for disruption/degeneration

Reverse-Engineering Strategy: Bridging Neurons to Behavior

A *= architecture class*

1. *"Circuit"*



3. *"Ecological niche/behavior"*

T *= task loss*

1. <u>Operationalize</u> a behavioral domain of interest.
Ways in which brains surpass current engineered systems? (e.g. performance, size, energy, etc)

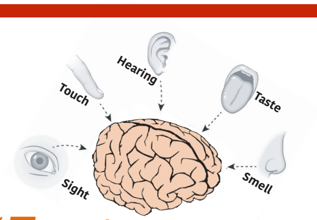
2. <u>Hypothesize</u> architectures and tasks (loss functions).
<u>Optimize</u> for task (evolution *in silico*).
<u>Predict</u> <u>held out</u> neural & behavioral data.

<u>3. Core Conceptual Insights:</u>
<u>Identify</u> the patterns of the best architectures & tasks.

A structural & functional *normative* understanding of the evolutionary constraints of the biological system to produce the behavior in (1).

Natural Science

Engineering



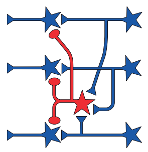
2. *"Environment"*

D *= data stream*

Reverse-Engineering Strategy: Bridging Neurons to Behavior

A = *architecture class*

1. *"Circuit"*



T = *task loss*

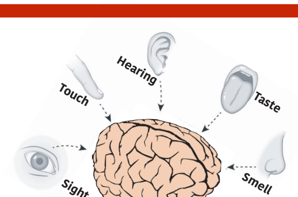
3. *"Ecological niche/behavior"*

1. Operationalize a behavioral domain of interest.
Ways in which brains surpass current engineered systems? (e.g. performance, size, energy, etc)

2. Hypothesize architectures and tasks (loss functions).
Optimize for task (evolution *in silico*).
Predict held out neural & behavioral data.

3. Core Conceptual Insights:
Identify the patterns of the best architectures & tasks.

A structural & functional *normative* understanding of the evolutionary constraints of the biological system to produce the behavior in (1).



2. *"Environment"*

D = *data stream*

Natural Science Engineering

Outline

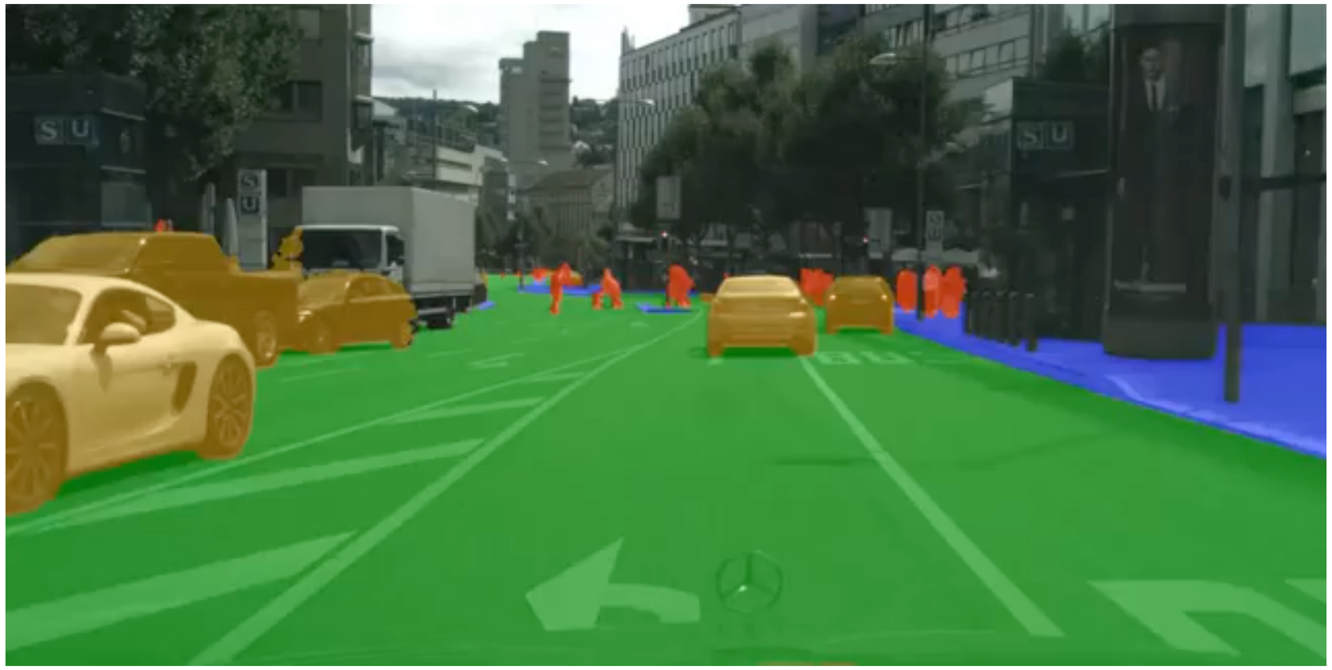
- Recurrent Connections in the Primate Ventral Stream
- Goal-Driven Models of Mouse Visual Cortex
- Heterogeneity in Rodent Medial Entorhinal Cortex
- Future Directions

Outline

- Recurrent Connections in the Primate Ventral Stream
- Goal-Driven Models of Mouse Visual Cortex
- Heterogeneity in Rodent Medial Entorhinal Cortex
- Future Directions

From Neurons to Behavior

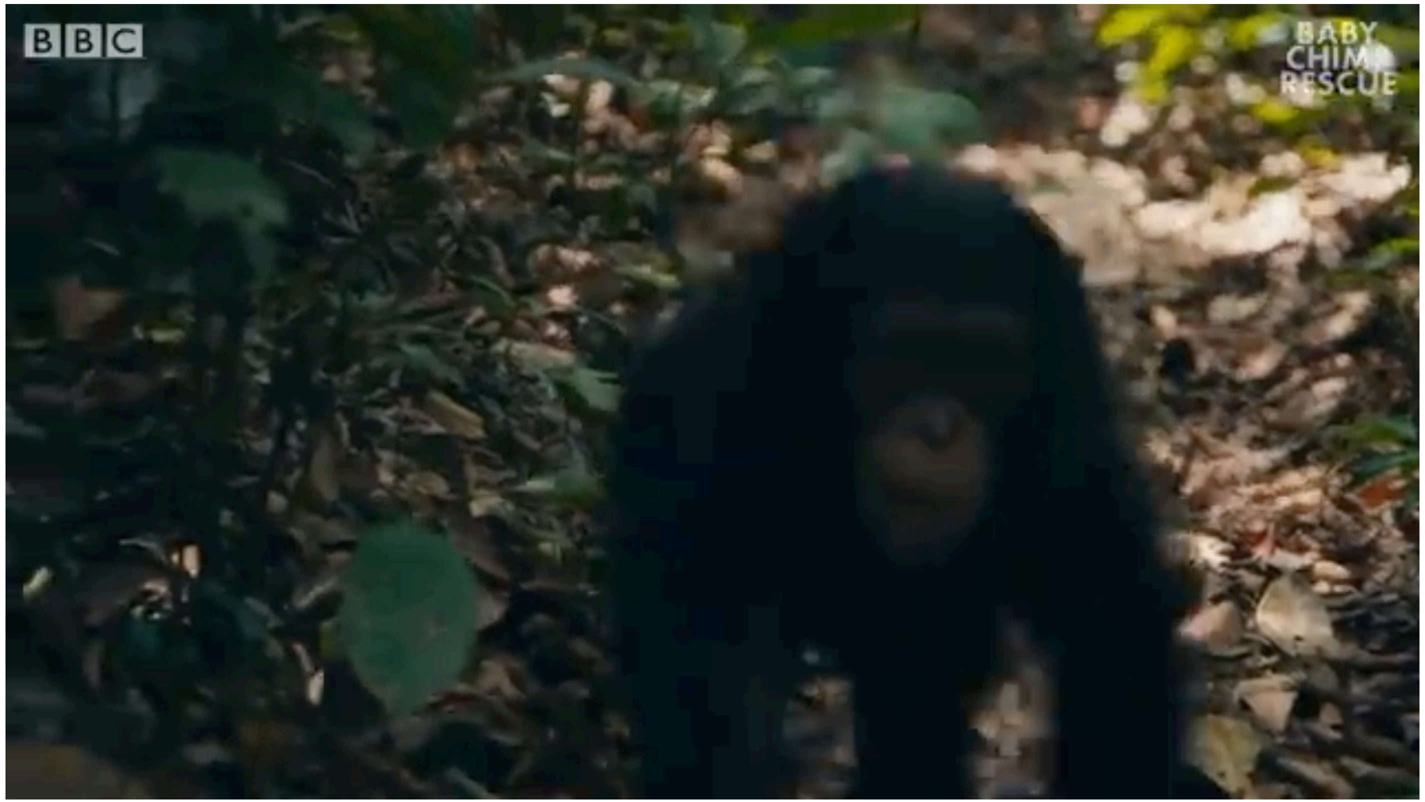
Scene Understanding



Multi-Step Planning

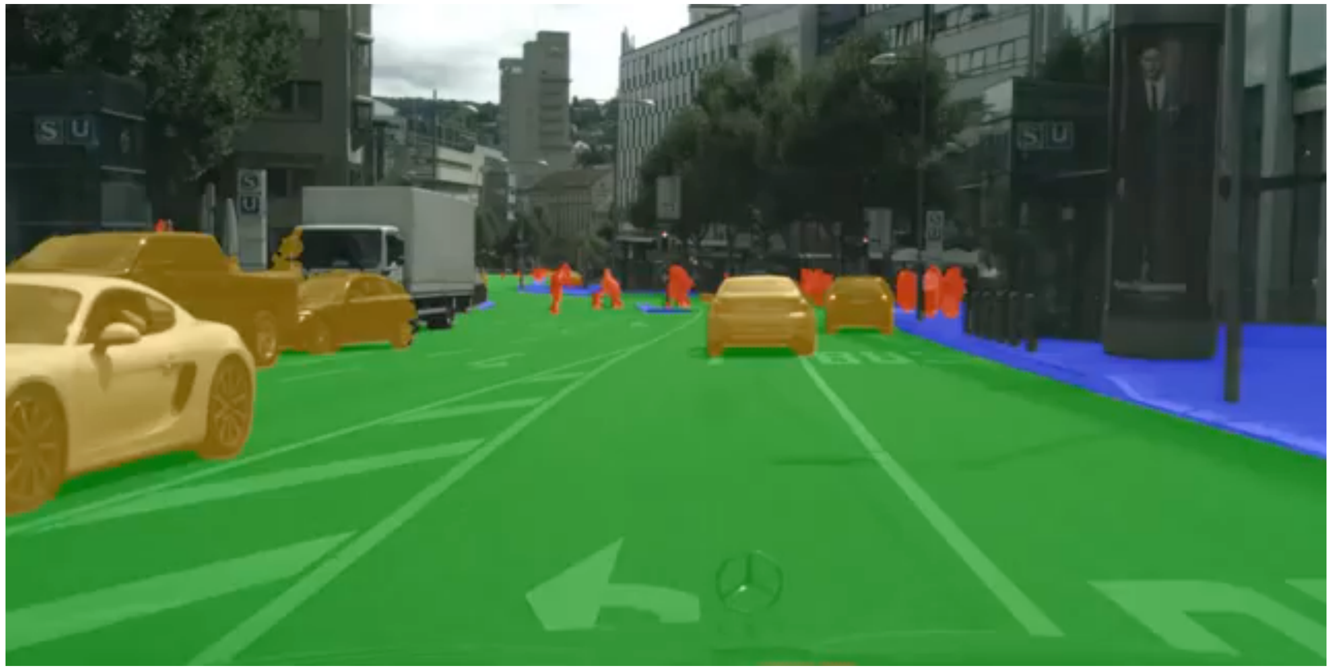
Navigation

Flexible Embodiment



From Neurons to Behavior

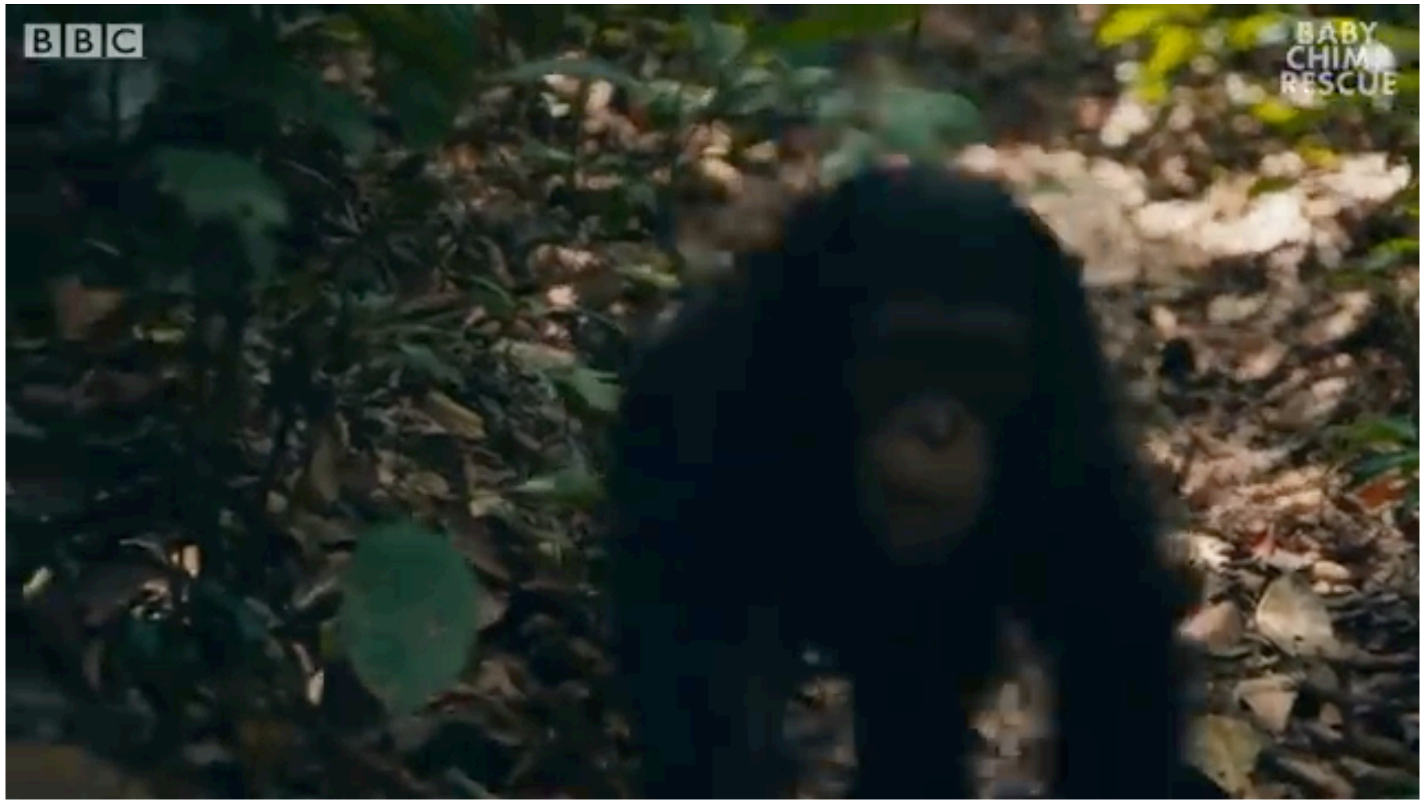
Scene Understanding



Multi-Step Planning

Navigation

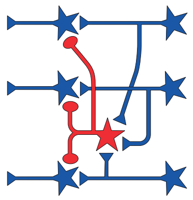
Flexible Embodiment



Recurrent Connections in the Primate Ventral Stream

A = *architecture class*

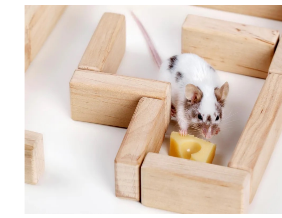
1. *"Circuit"*



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

T = *task loss*

3. *"Ecological niche/behavior"*



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

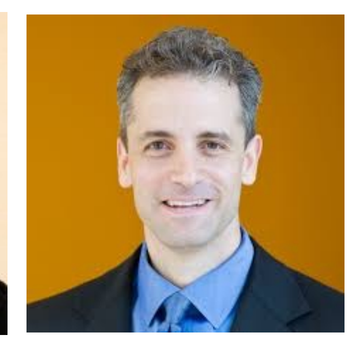
Jonas Kubilius

Kohitij Kar

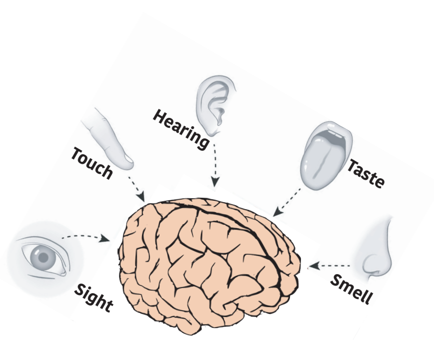
Surya Ganguli

Javier Sagastuy

David Sussillo



Jim DiCarlo



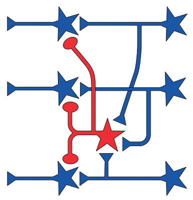
2. *"Environment"*

D = *data stream*

Recurrent Connections in the Primate Ventral Stream

A = *architecture class*

1. *"Circuit"*



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

T = *task loss*

3. *"Ecological niche/behavior"*

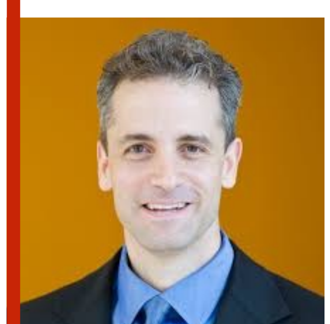
A. Nayebi, *et al.*
Recurrent Connections in the Primate Ventral Visual Stream Mediate a Tradeoff Between Task Performance and Network Size During Core Object Recognition. *Neural Computation 2022*

Daniel Yamins

Daniel Bear

Jonas Kubilius

Kohitij Kar

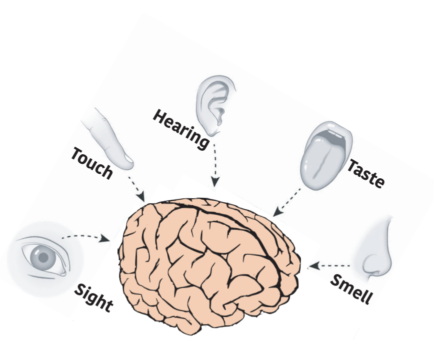


Surya Ganguli

Javier Sagastuy

David Sussillo

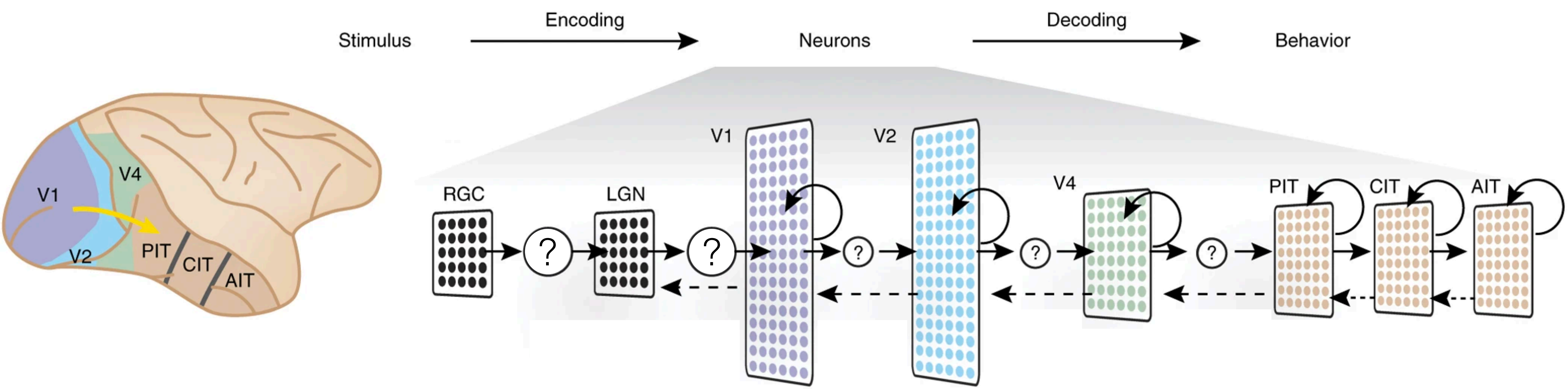
Jim DiCarlo



2. *"Environment"*

D = *data stream*

CNNs as Models of Object Recognition



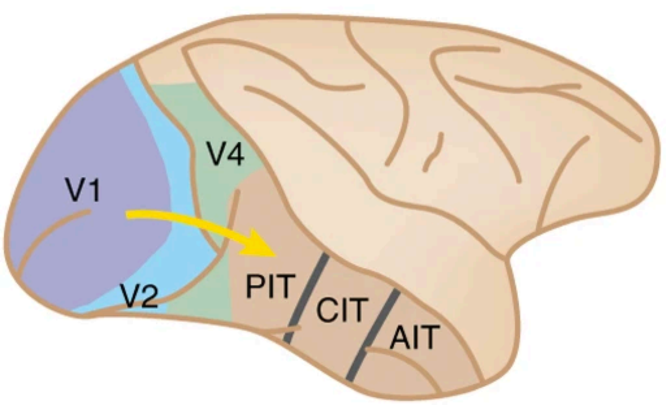
Convolutional Neural Networks (CNNs)

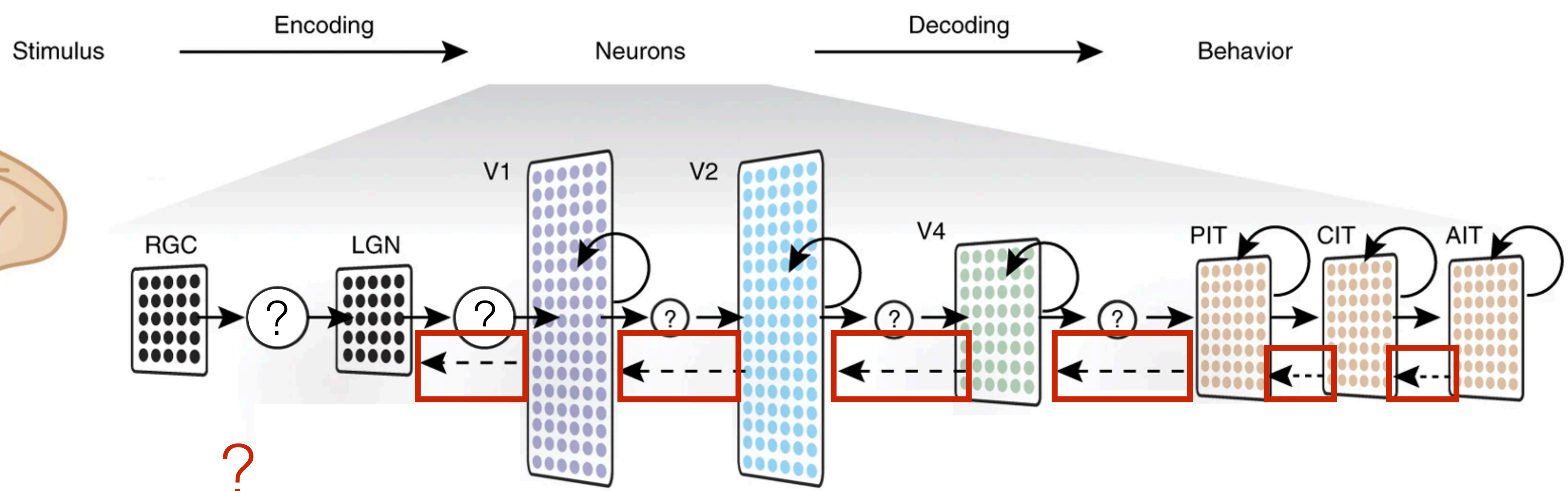
Fukushima, 1979; Lecun, 1995

CNNs are inspired by visual neuroscience:

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

…but lack feedback connections





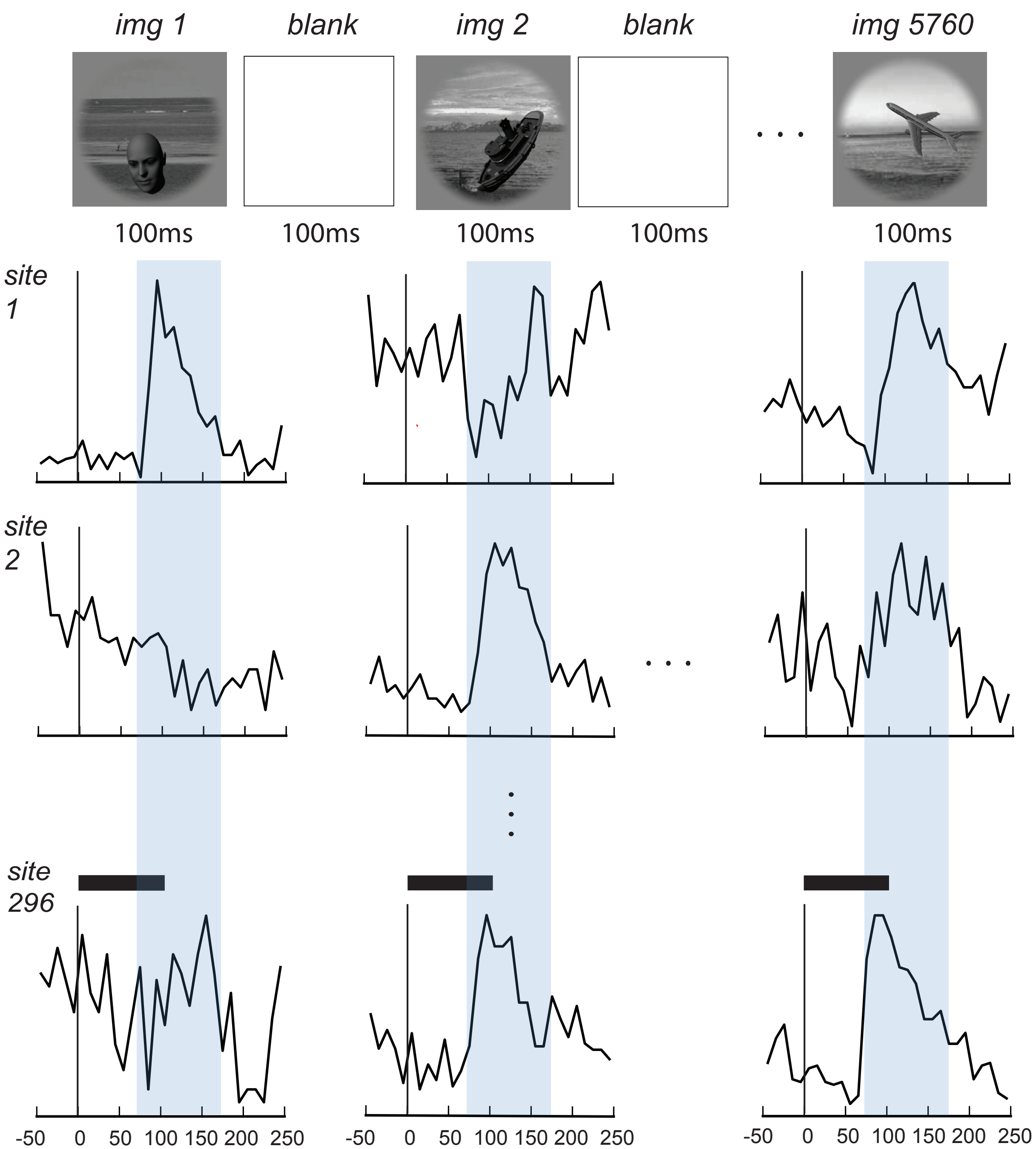
Convolutional Neural Networks (CNNs)

Fukushima, 1979; Lecun, 1995

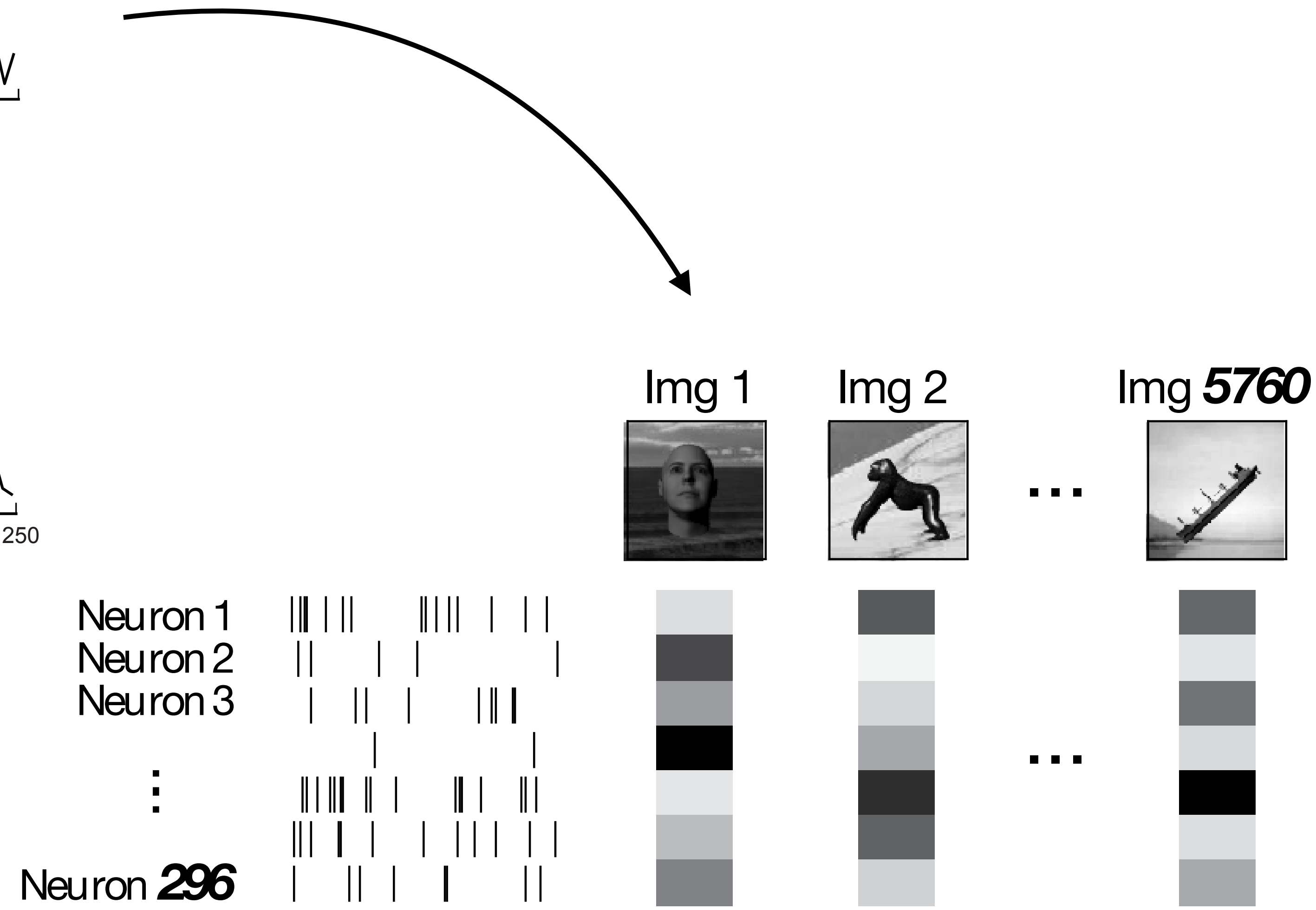
CNNs are inspired by visual neuroscience:

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

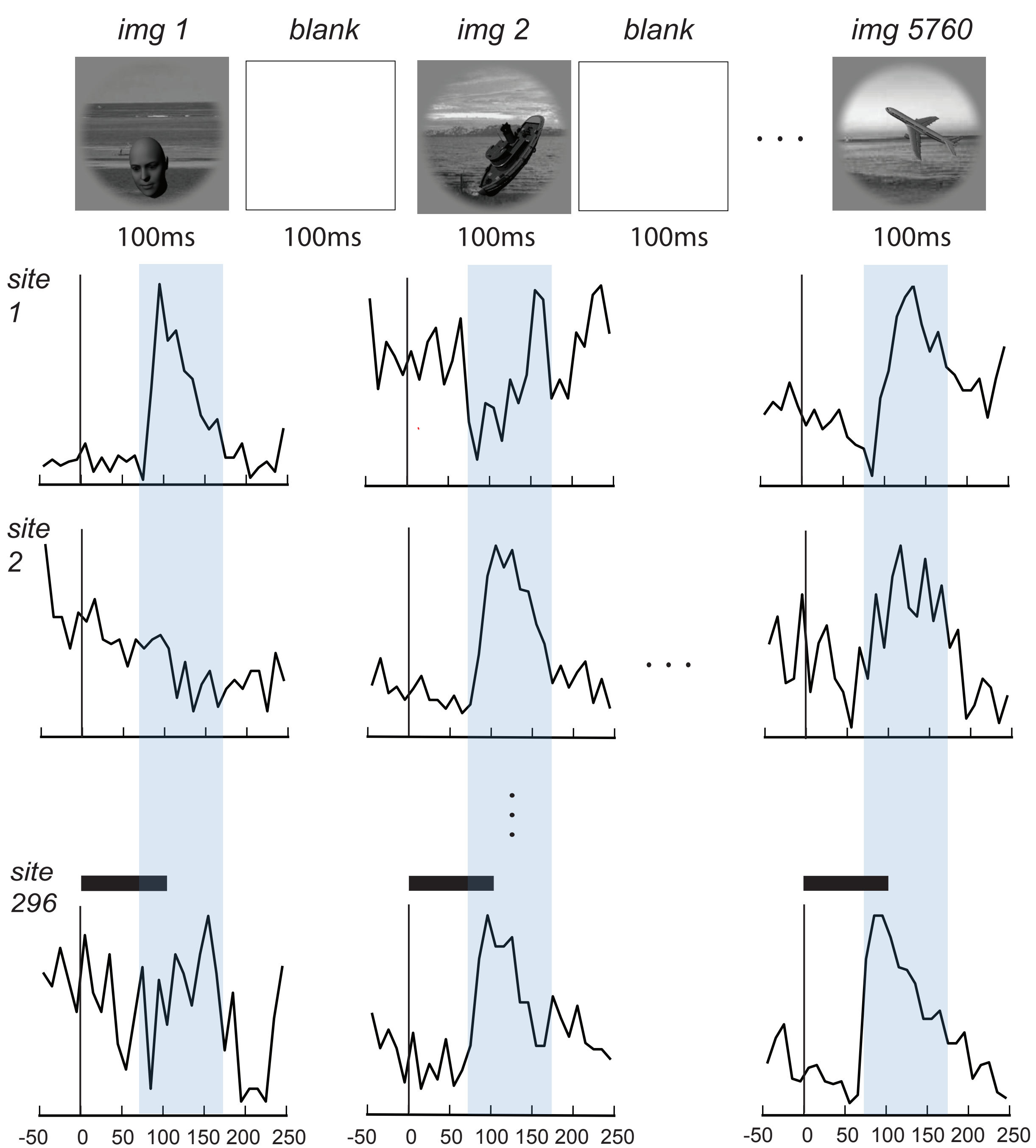
So far, only explaining temporal average of responses



e.g. Binned spike counts 70ms-170ms post stimulus presentation

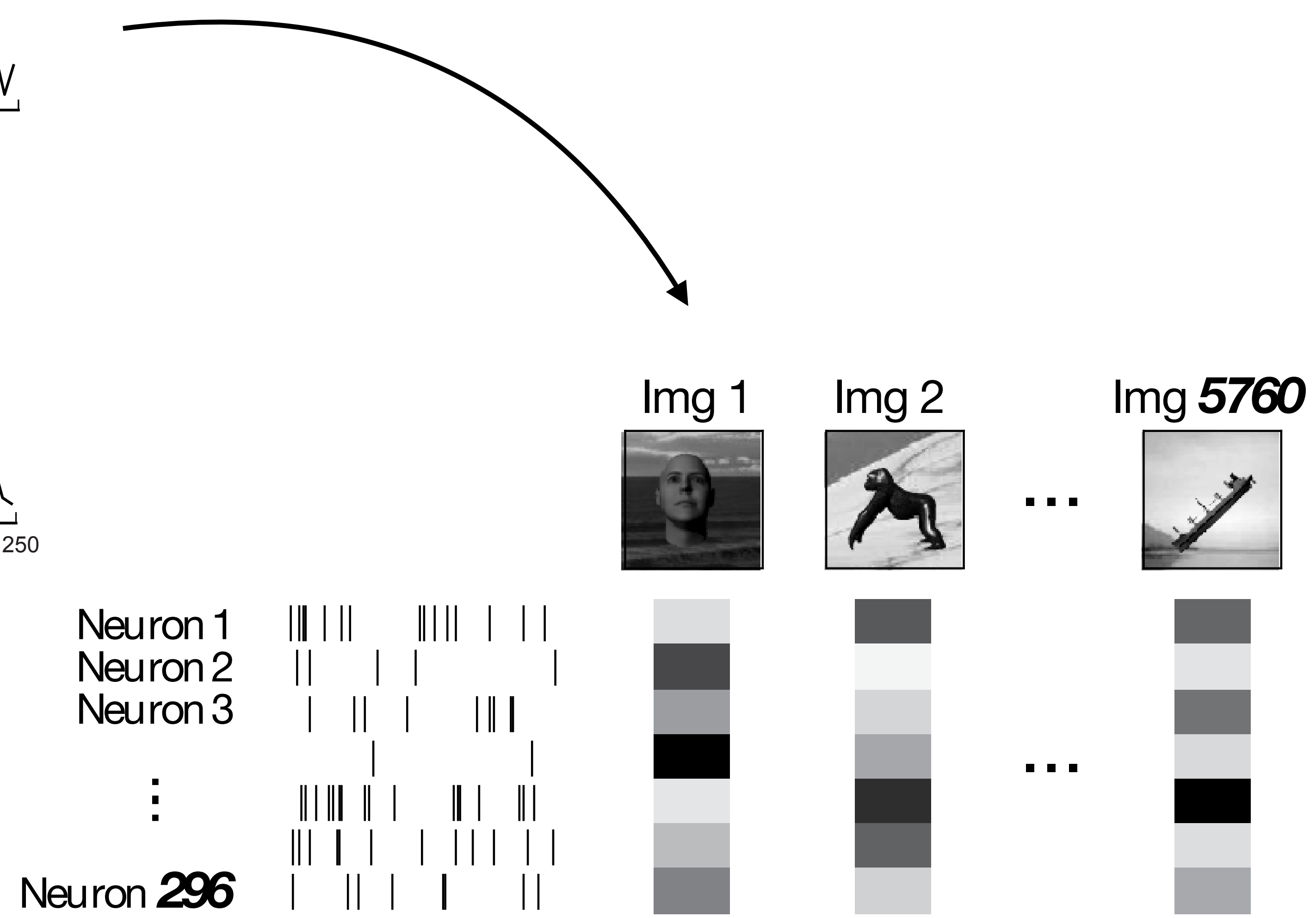


So far, only explaining temporal average of responses



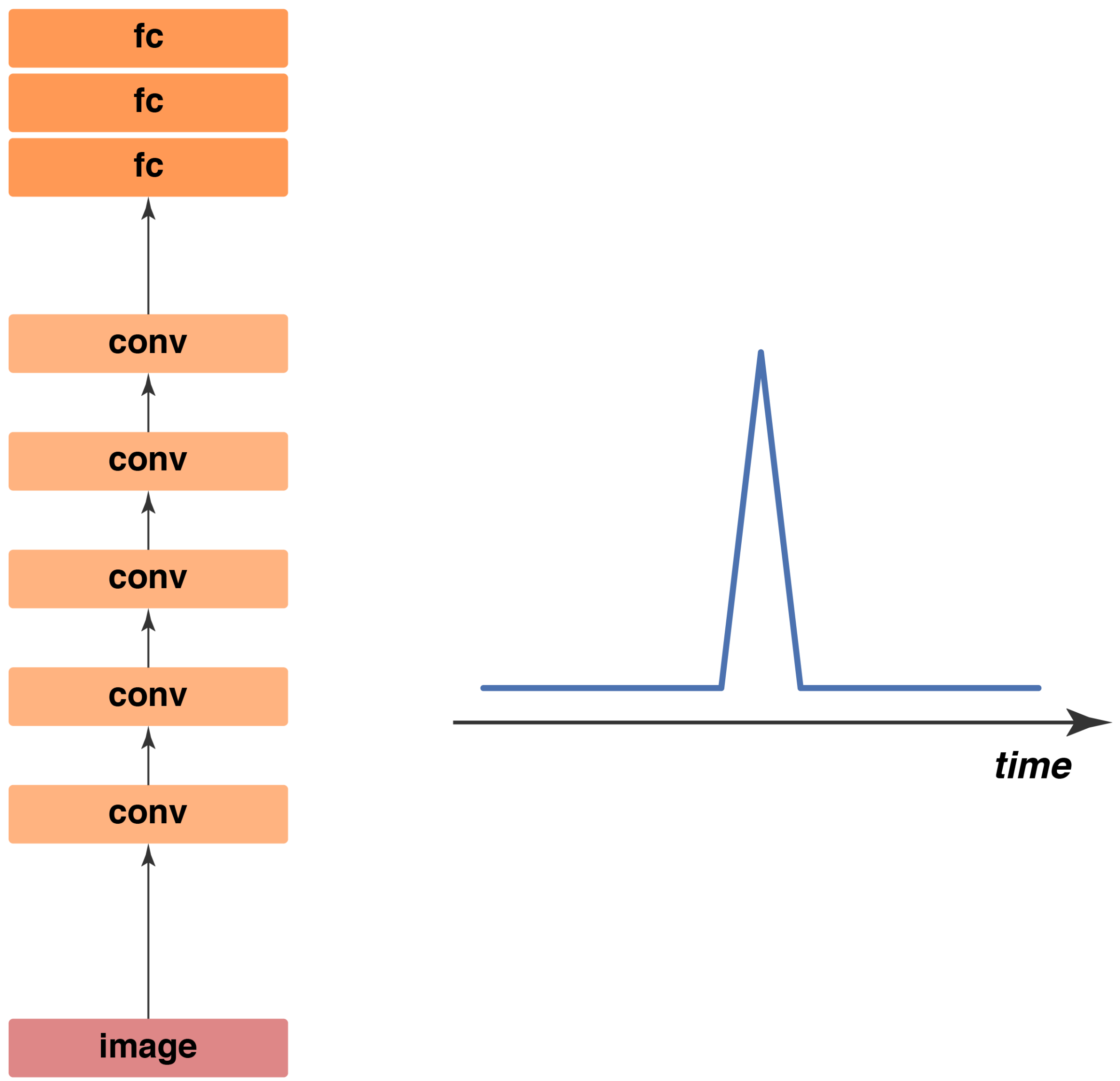
e.g. Binned spike counts 70ms-170ms post stimulus presentation

but actually the data is highly reliable at much finer grain — 10ms bins



Trajectory possibilities due to recurrent connections

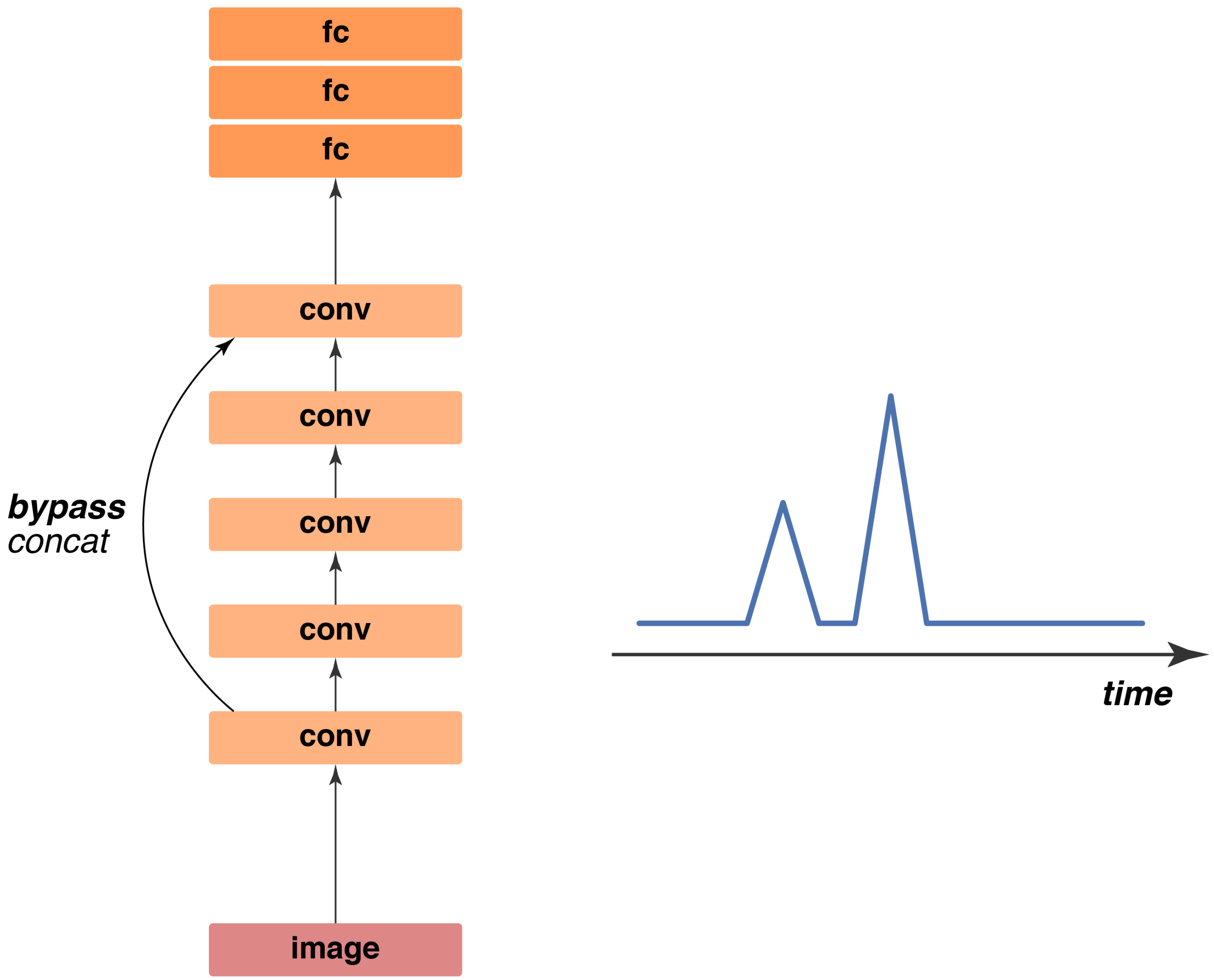
Simple feedforward networks simple dynamics:



courtesy Jonas Kubilius

Trajectory possibilities due to recurrent connections

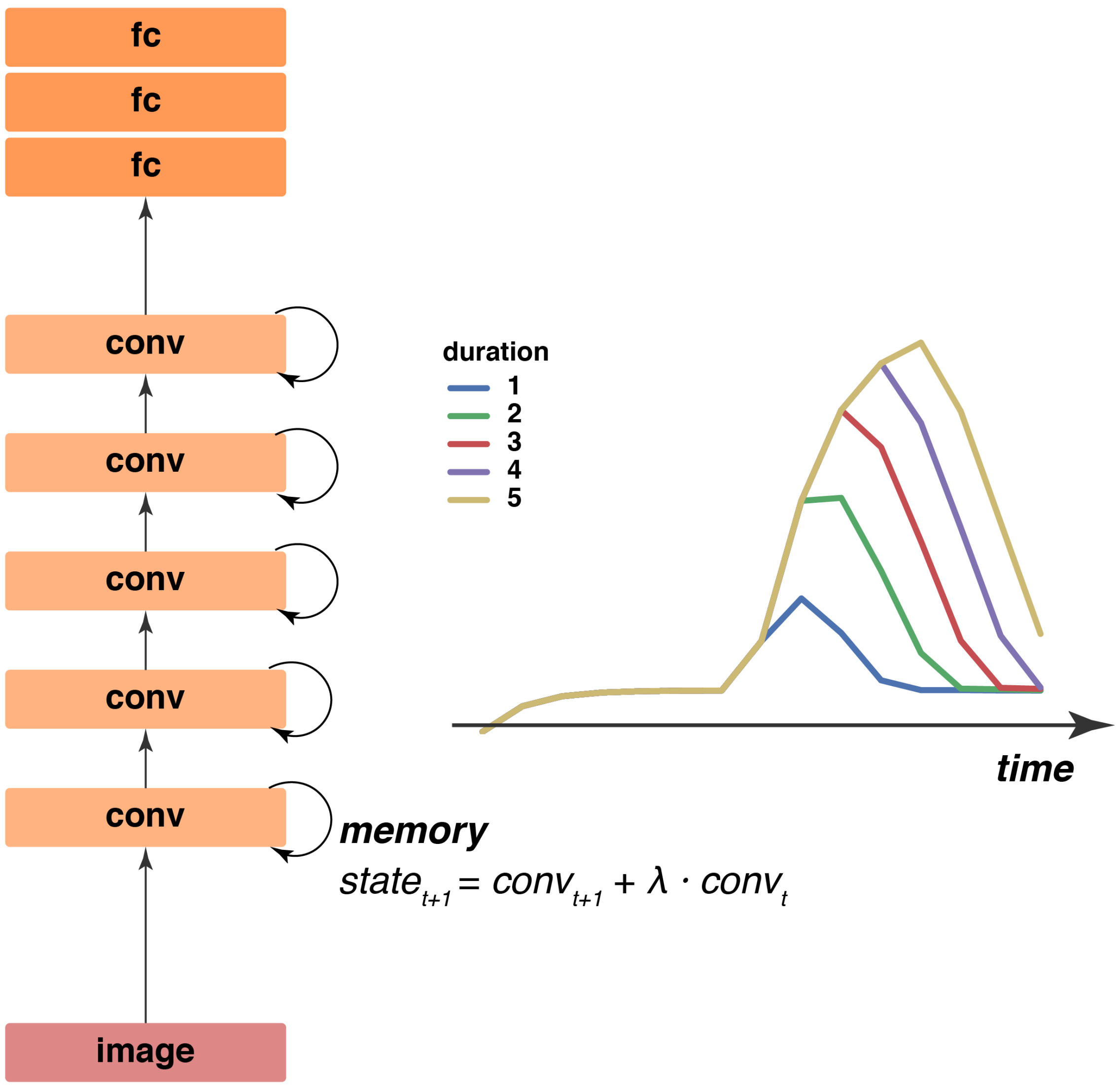
Dynamics more interesting with bypasses:



courtesy Jonas Kubilius

Trajectory possibilities due to recurrent connections

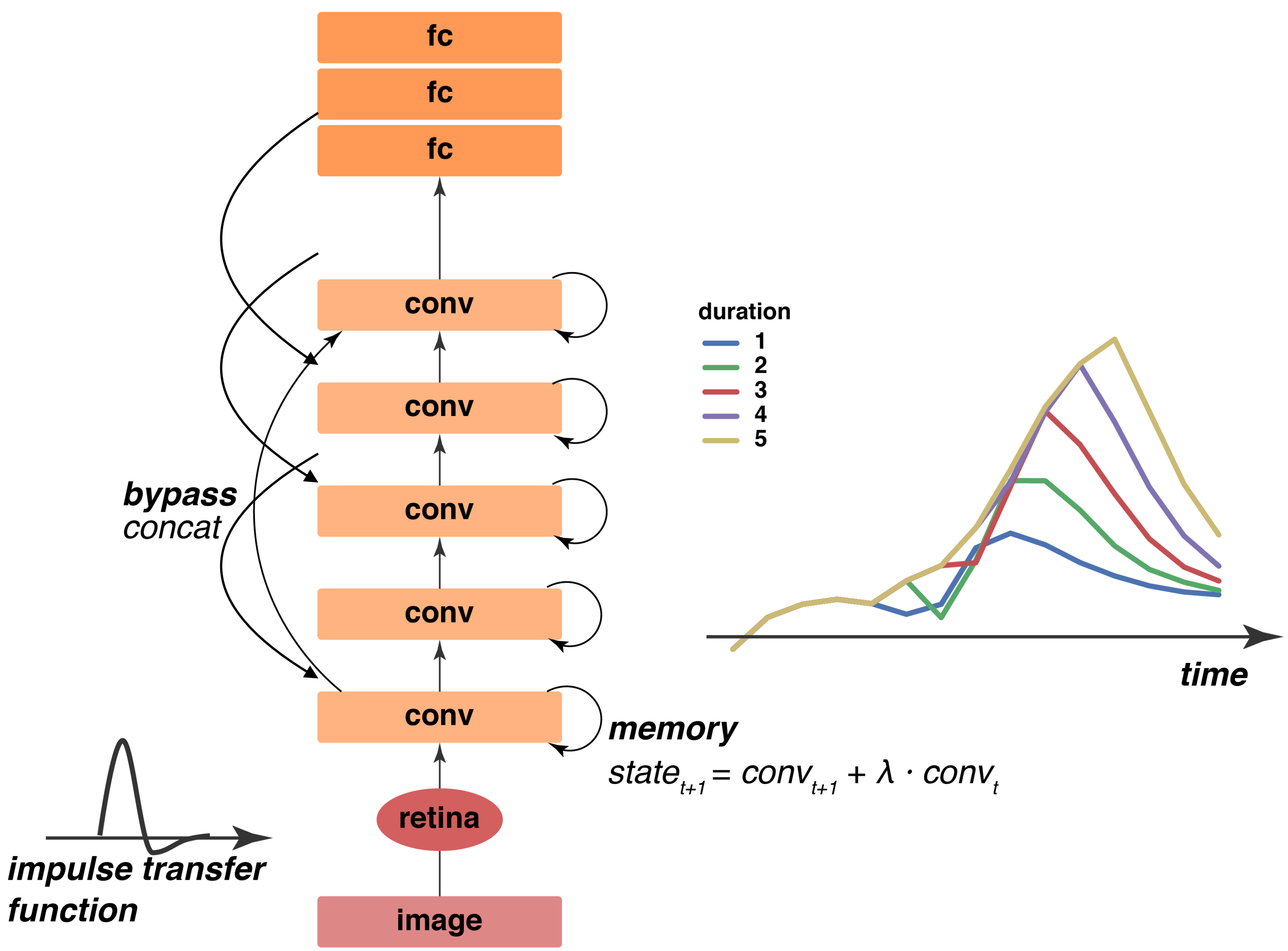
Dynamics more interesting with bypasses, local recurrence:



courtesy Jonas Kubilius

Trajectory possibilities due to recurrent connections

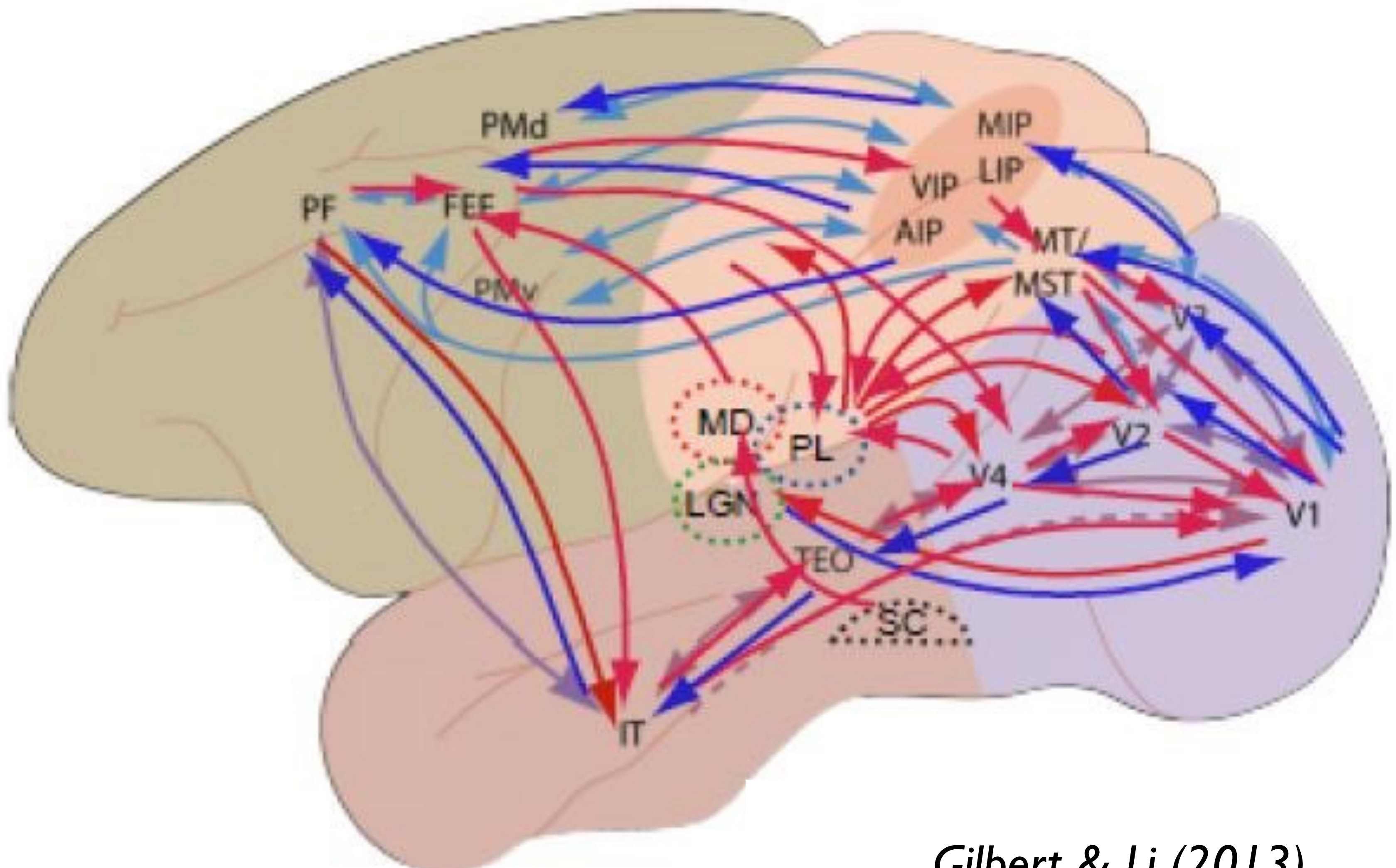
Dynamics more interesting with bypasses, local recurrence, long-range feedback:



courtesy Jonas Kubilius

Recurrent connections are ubiquitous

Recurrent connections are everywhere anatomically:



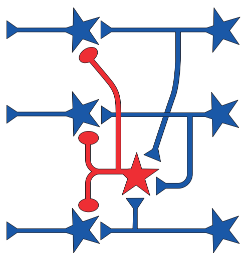
Gilbert & Li (2013)

. . . but what are they for?

Recurrent Connections in the Primate Ventral Stream

A = *architecture class*

1. *"Circuit"*

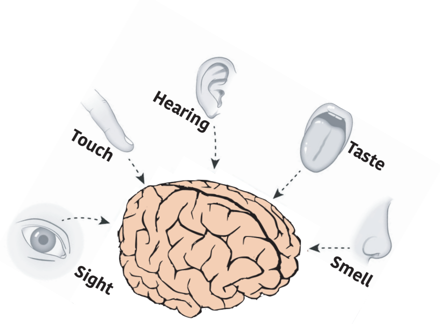


T = *task loss*

3. *"Ecological niche/behavior"*

<u>Neurobiological Puzzle:</u>

What is the role of recurrence in the primate ventral stream?



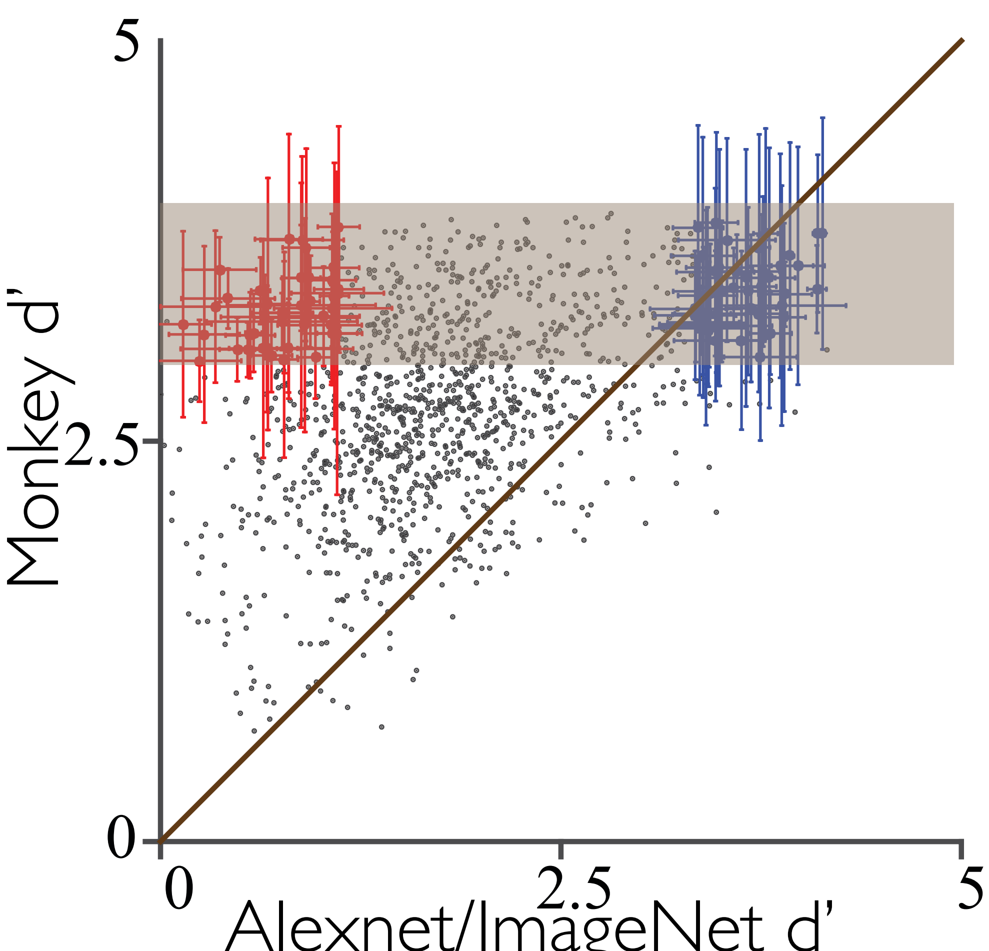
2. *"Environment"*

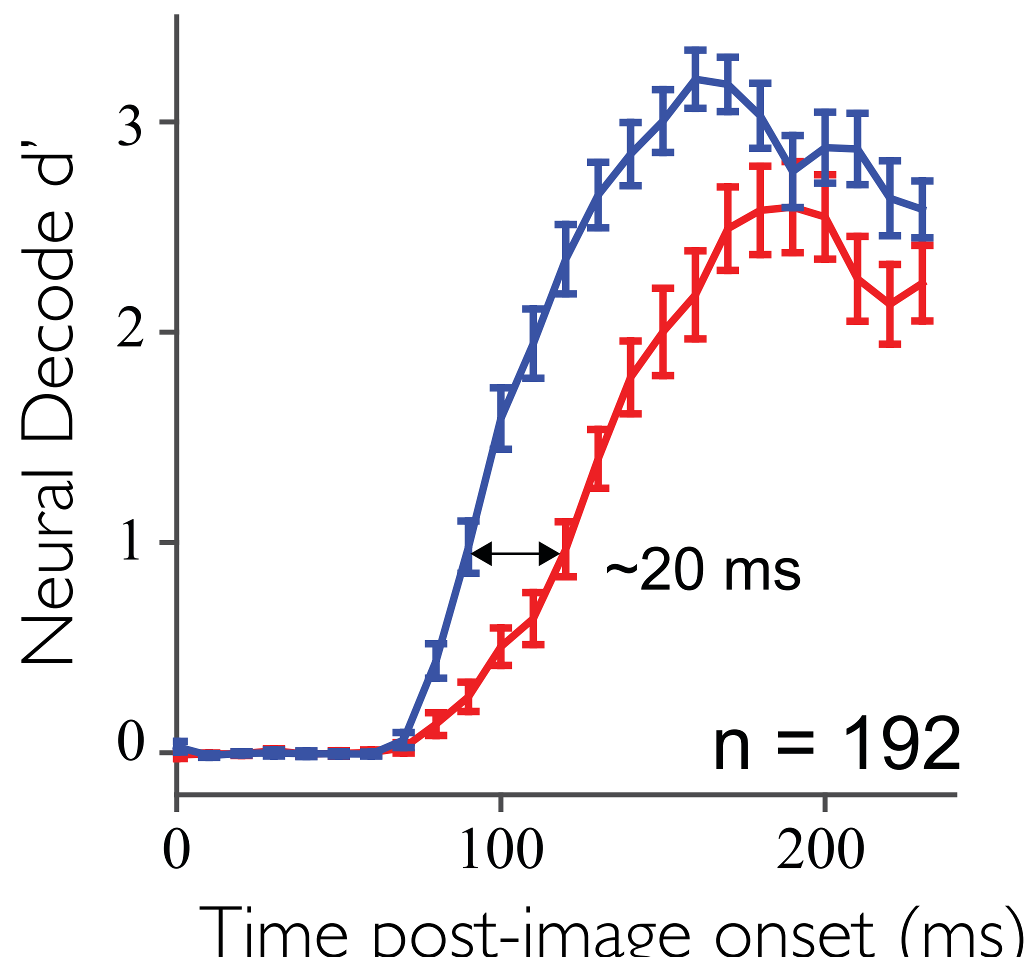
D = *data stream*

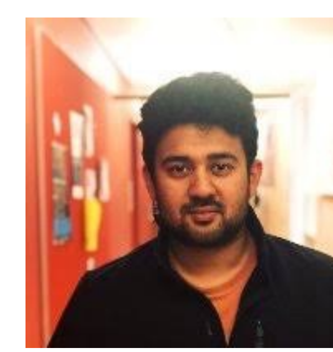
Evidence of Functional Relevance during Core Object Recognition

CNN-not-solved images ***are*** **solved by the ventral visual stream**

… even in range before 250ms (rough saccade time)





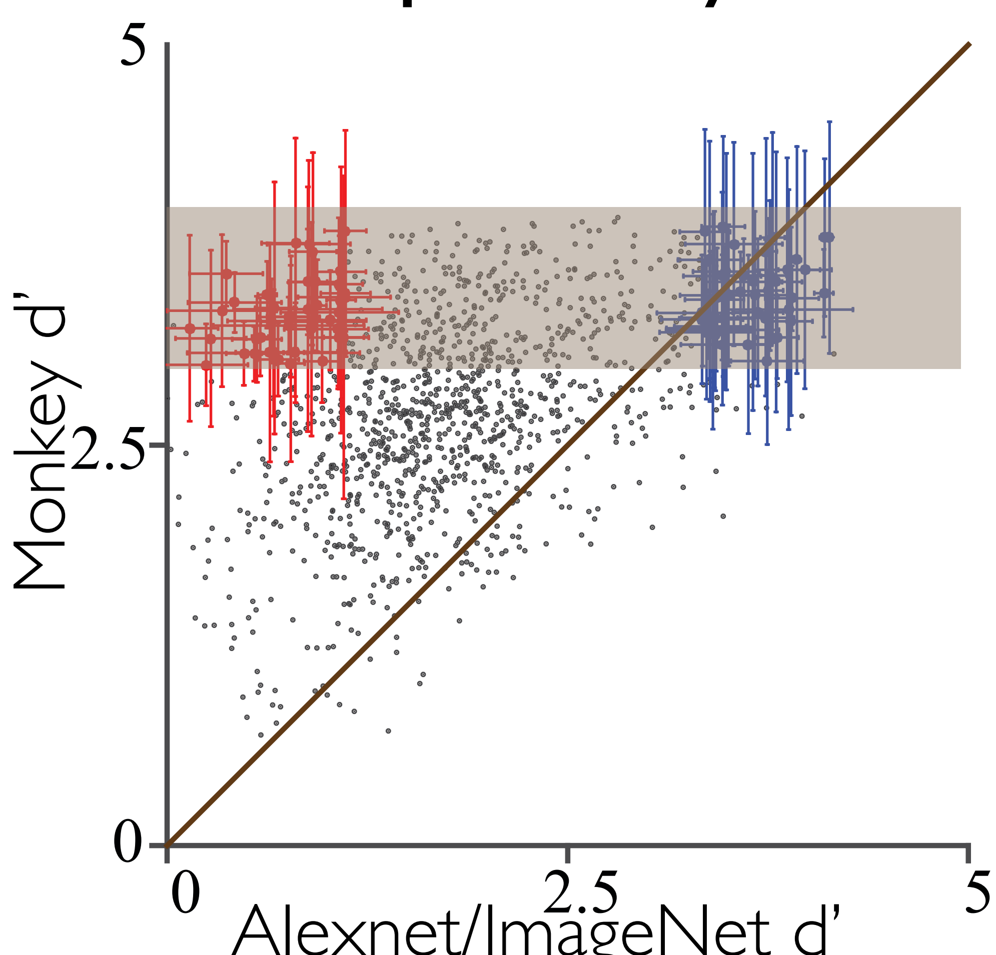


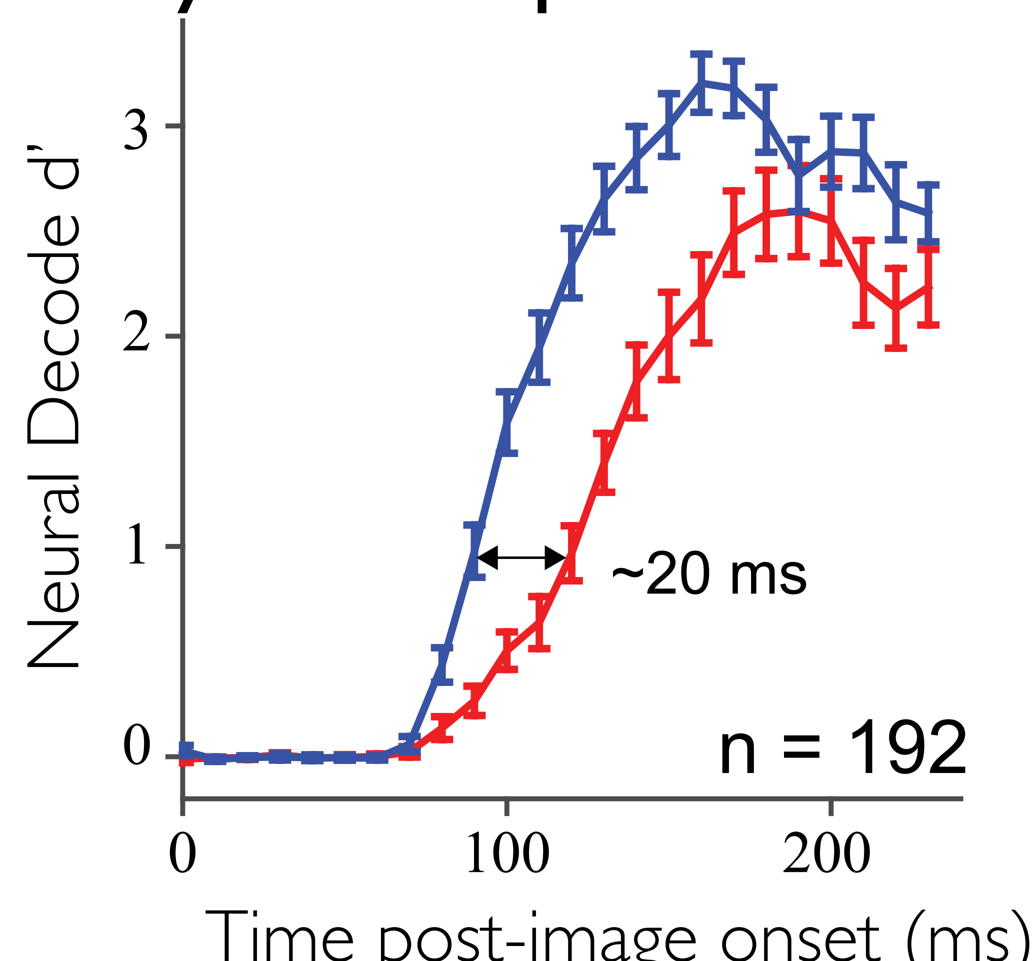
Kar et. al. (2019)

Evidence of Functional Relevance during Core Object Recognition

CNN-not-solved images <u>*are*</u> solved by the ventral visual stream

Can we explain *why* we see these dynamical patterns emerge in IT?



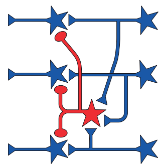


Kar et. al. (2019)

Convolutional Recurrent Networks (ConvRNNs)

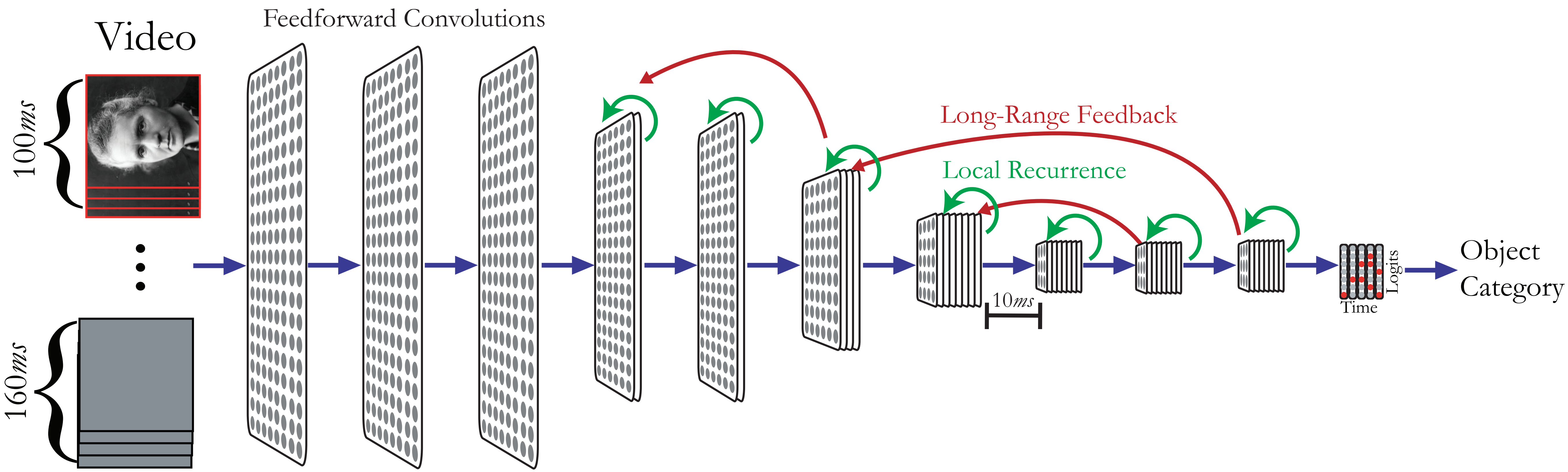
***A* = architecture class**

1. *"Circuit"*



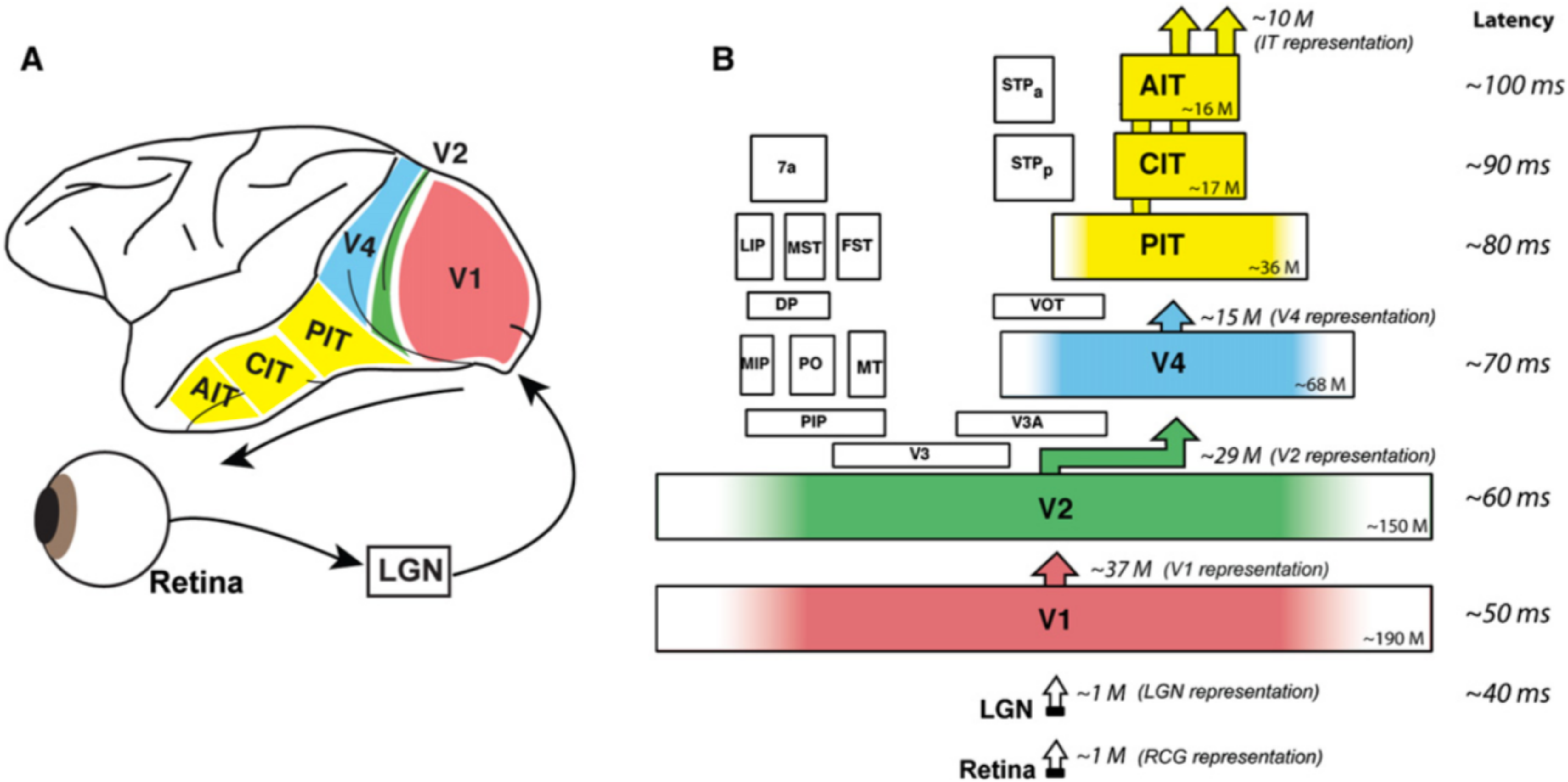
ConvRNNs

~~**CNNs**~~



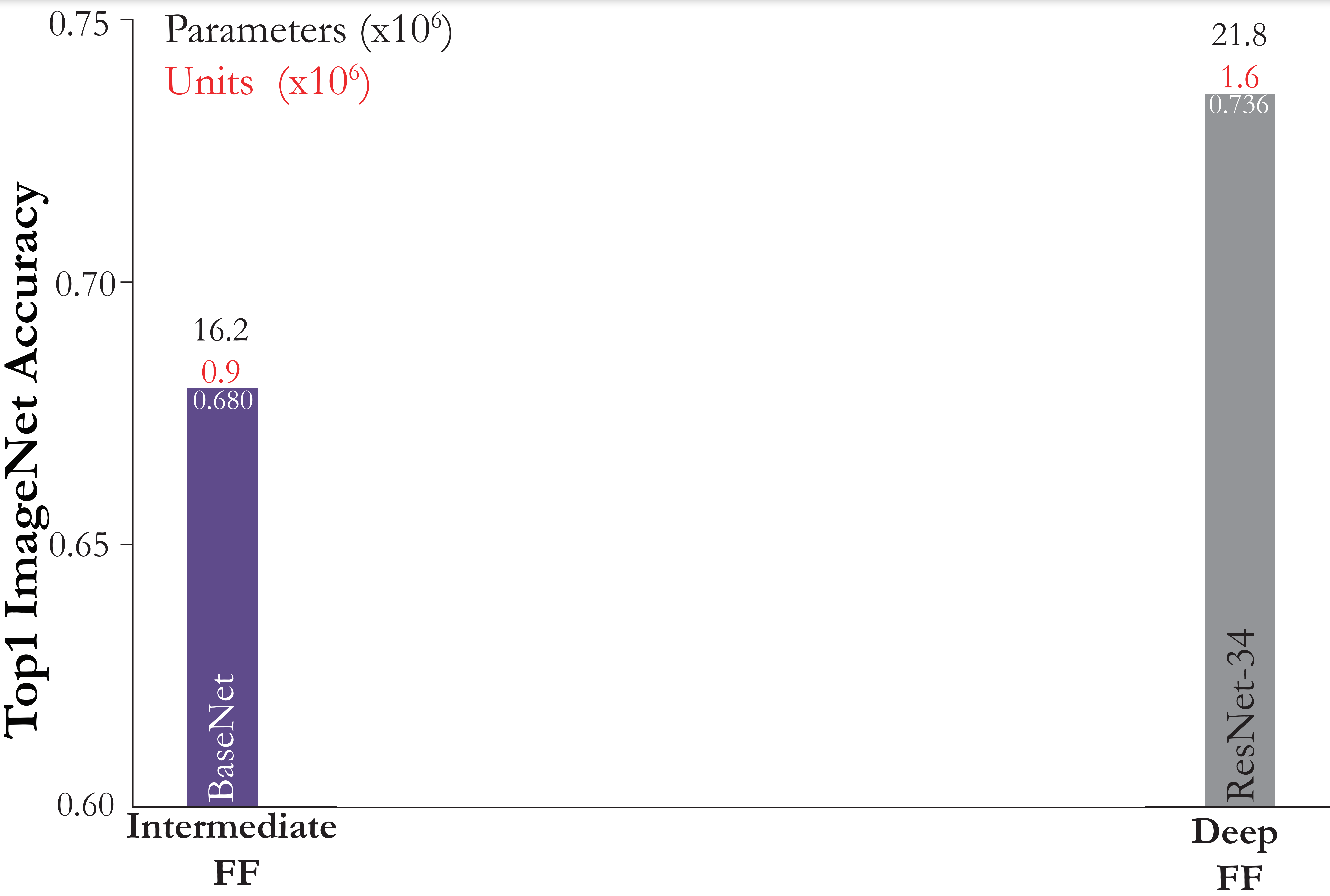
Each time-step (10 ms) is treated equally
— including feedforward steps

~10-12 "Layers" Plausible based on anatomy and timing



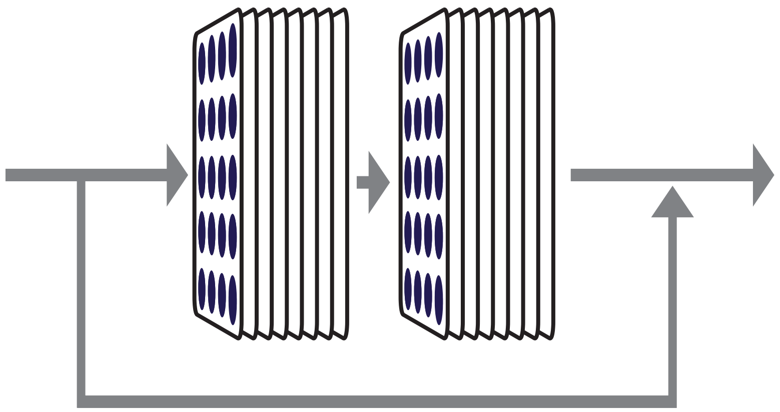
DiCarlo, Zoccolan, & Rust 2012

…But, such networks are not the most performant

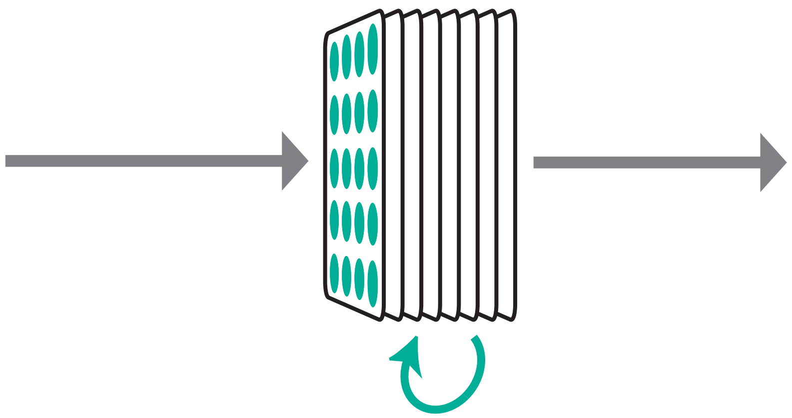


Many Choices of Local Recurrence

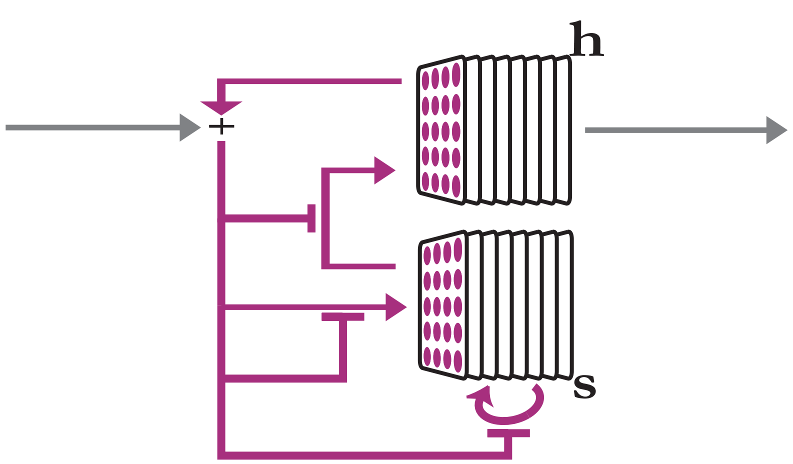
ResNet



SimpleRNN

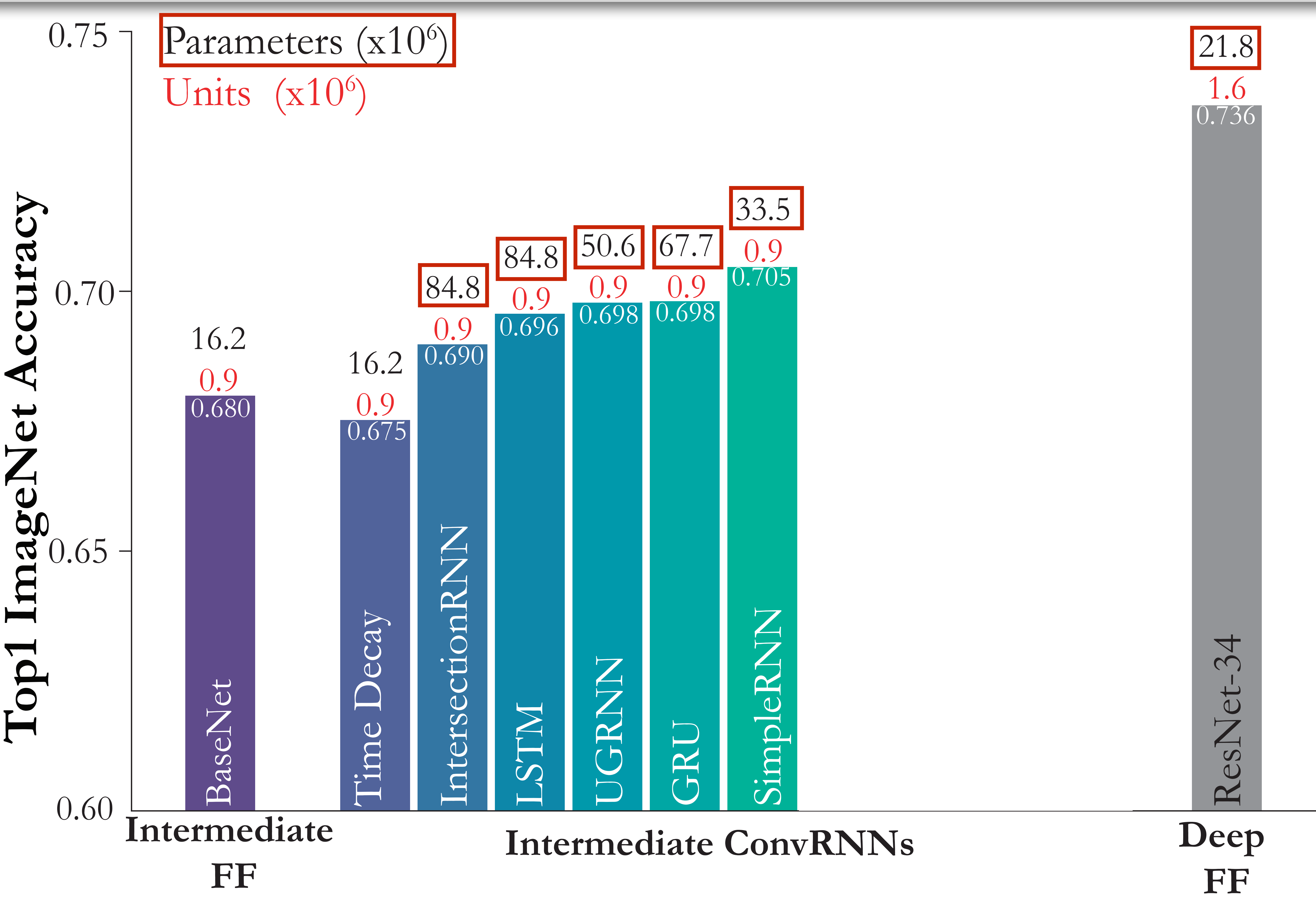


LSTM

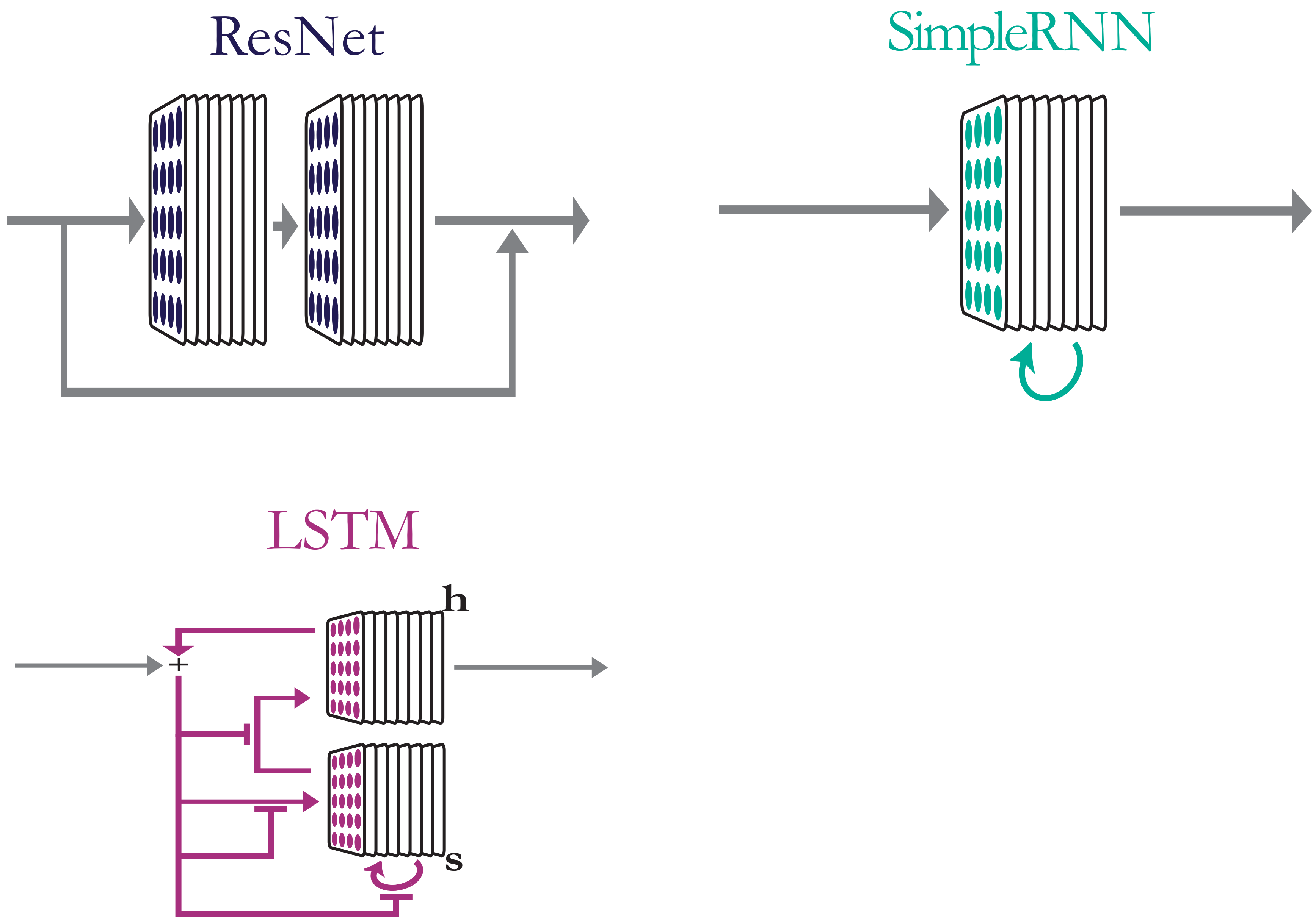


Circuit Diagram

Adding these helps somewhat, but add lots of parameters!

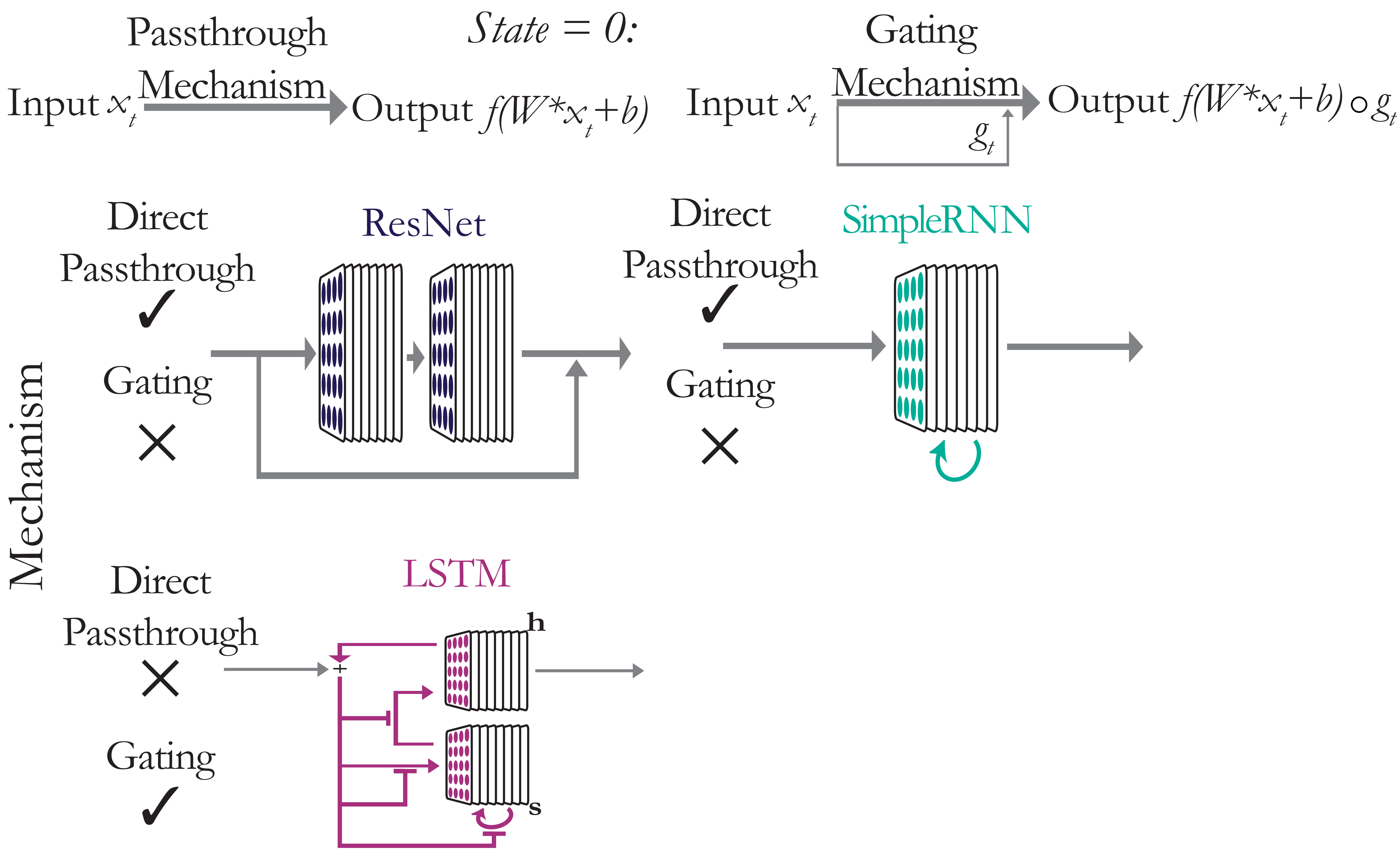


Many Choices of Local Recurrence



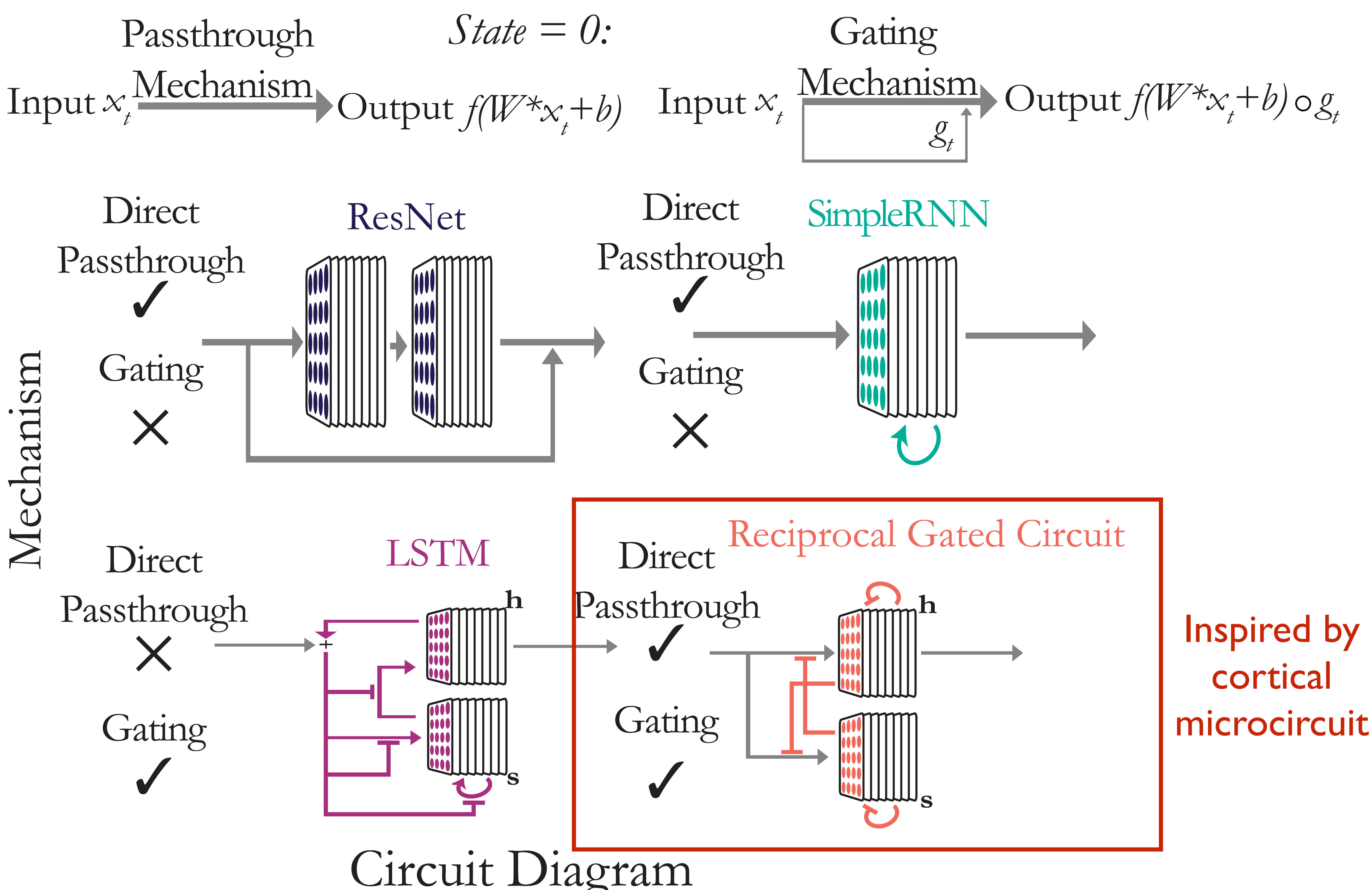
Circuit Diagram

Principles of Local Recurrence

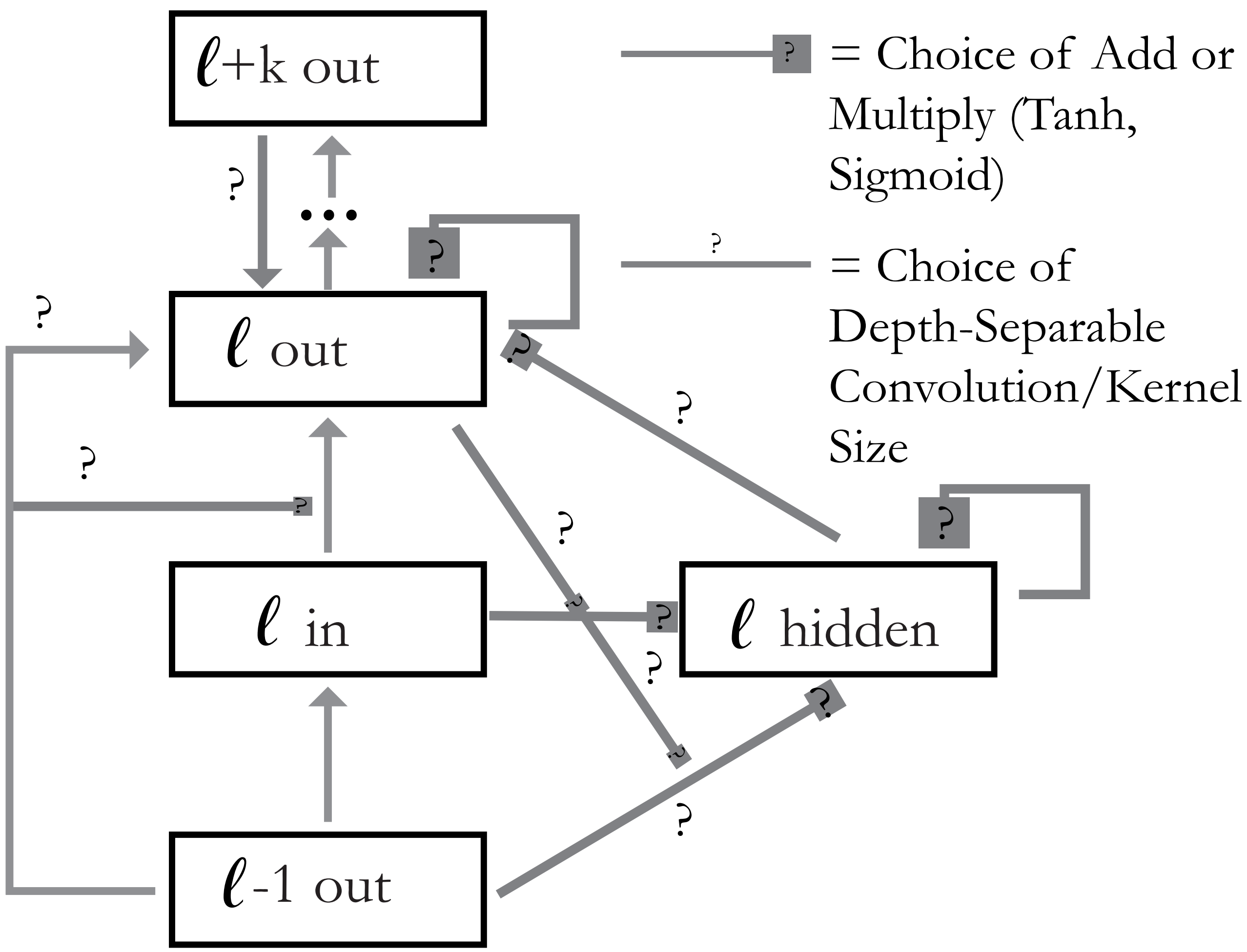


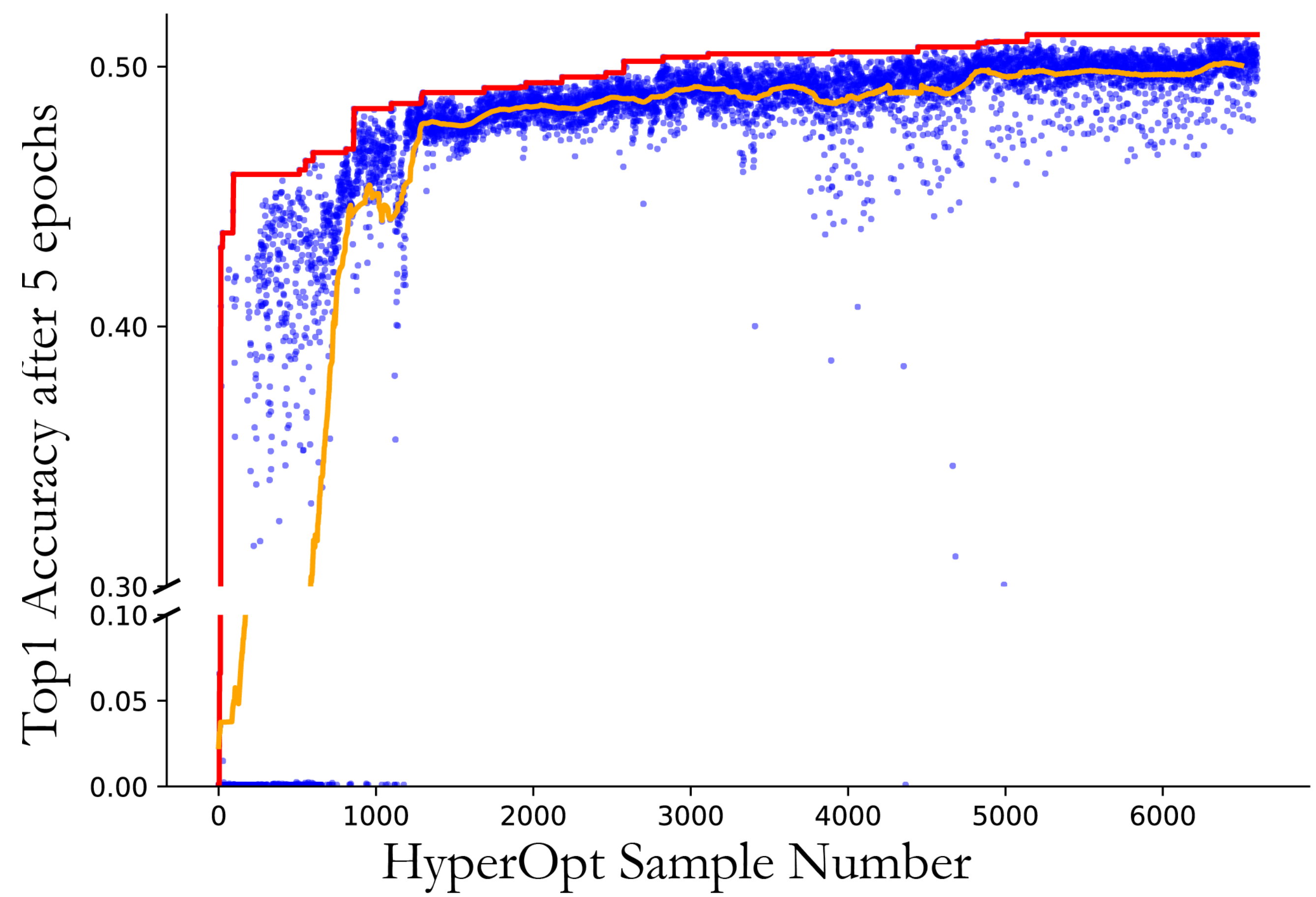
Circuit Diagram

Principles of Local Recurrence: Strong Circuit Constraints

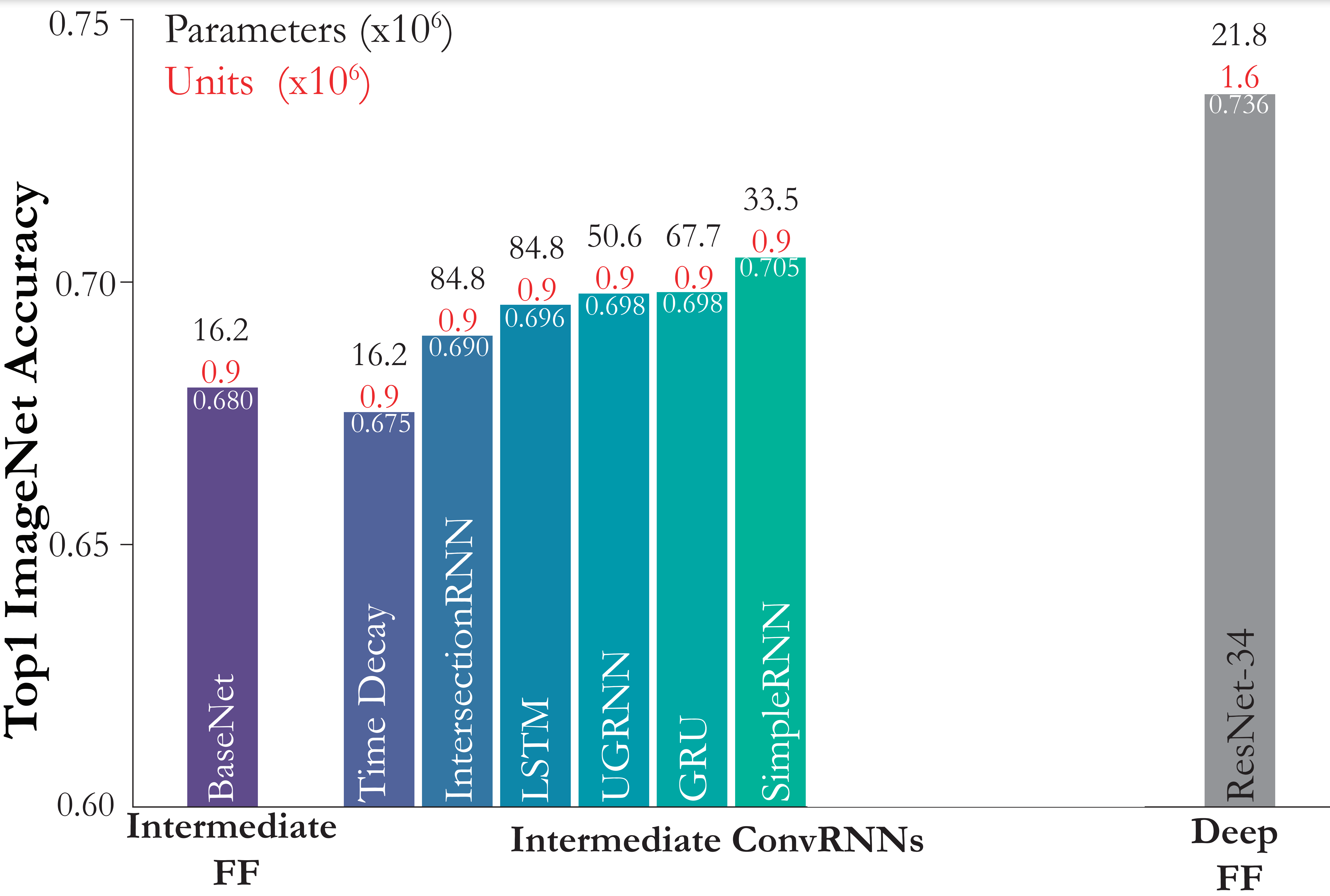


Search Over Local and Global Recurrence

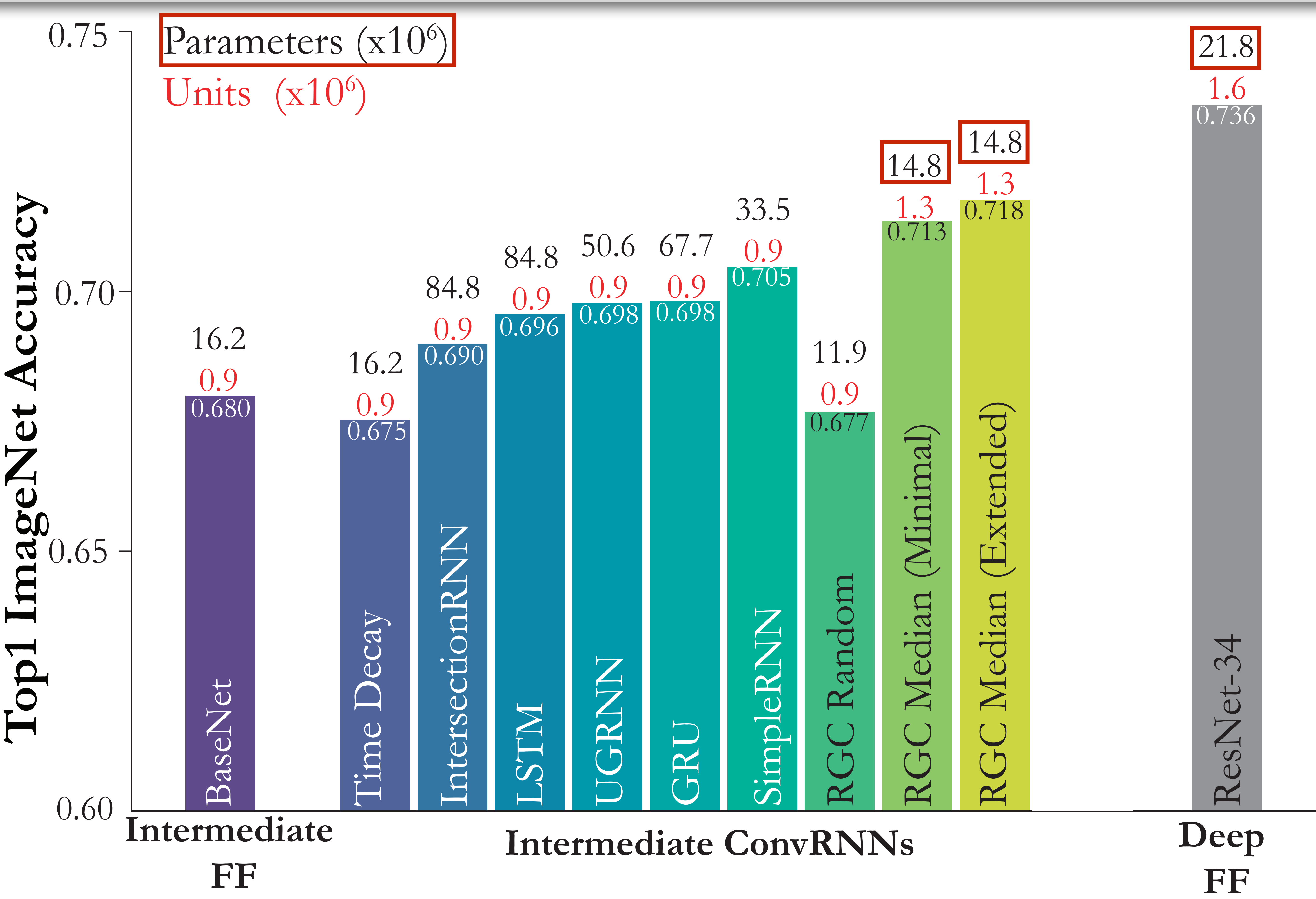




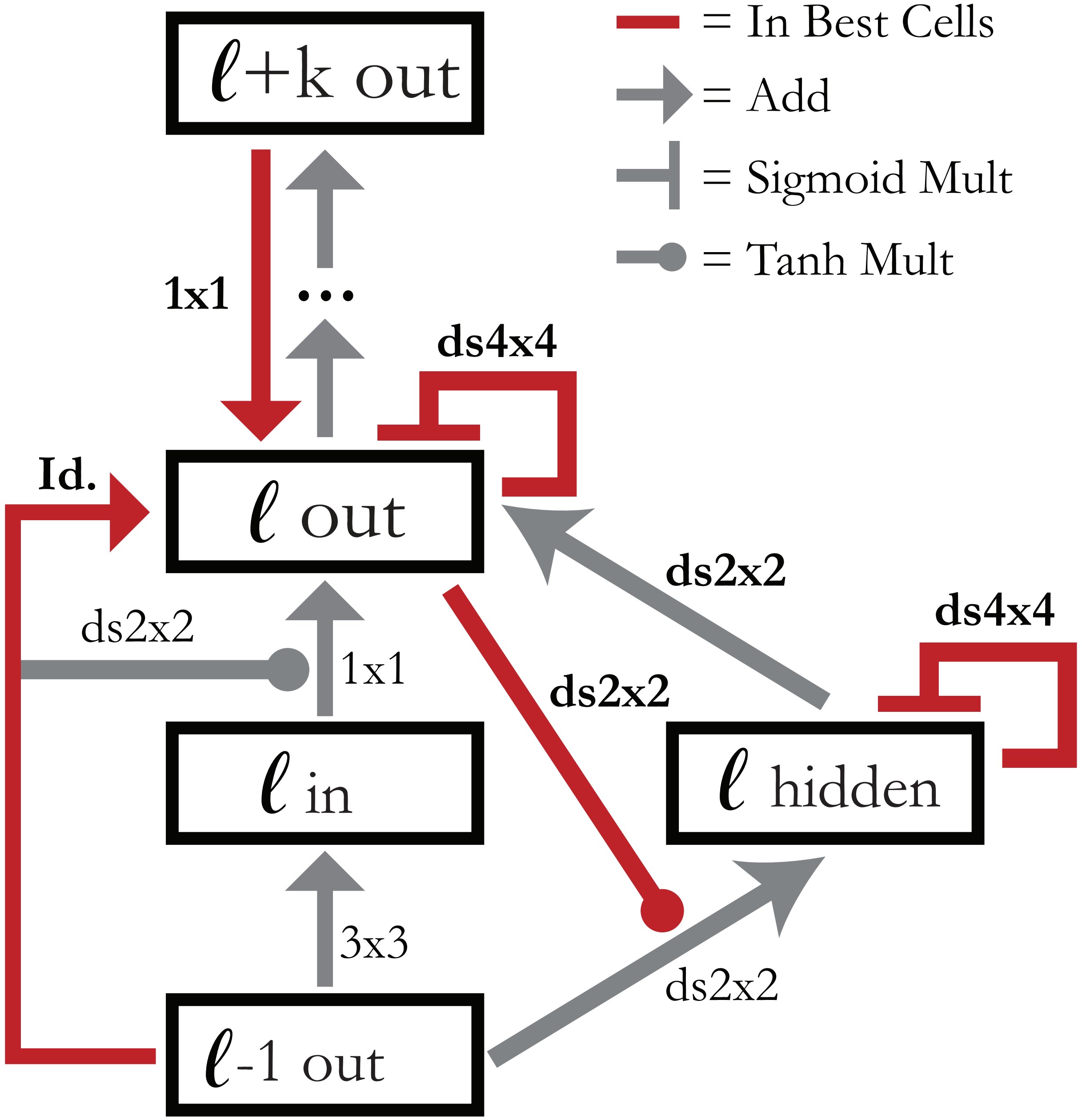
Evolutionary search yields improved performance



Evolutionary search yields improved performance

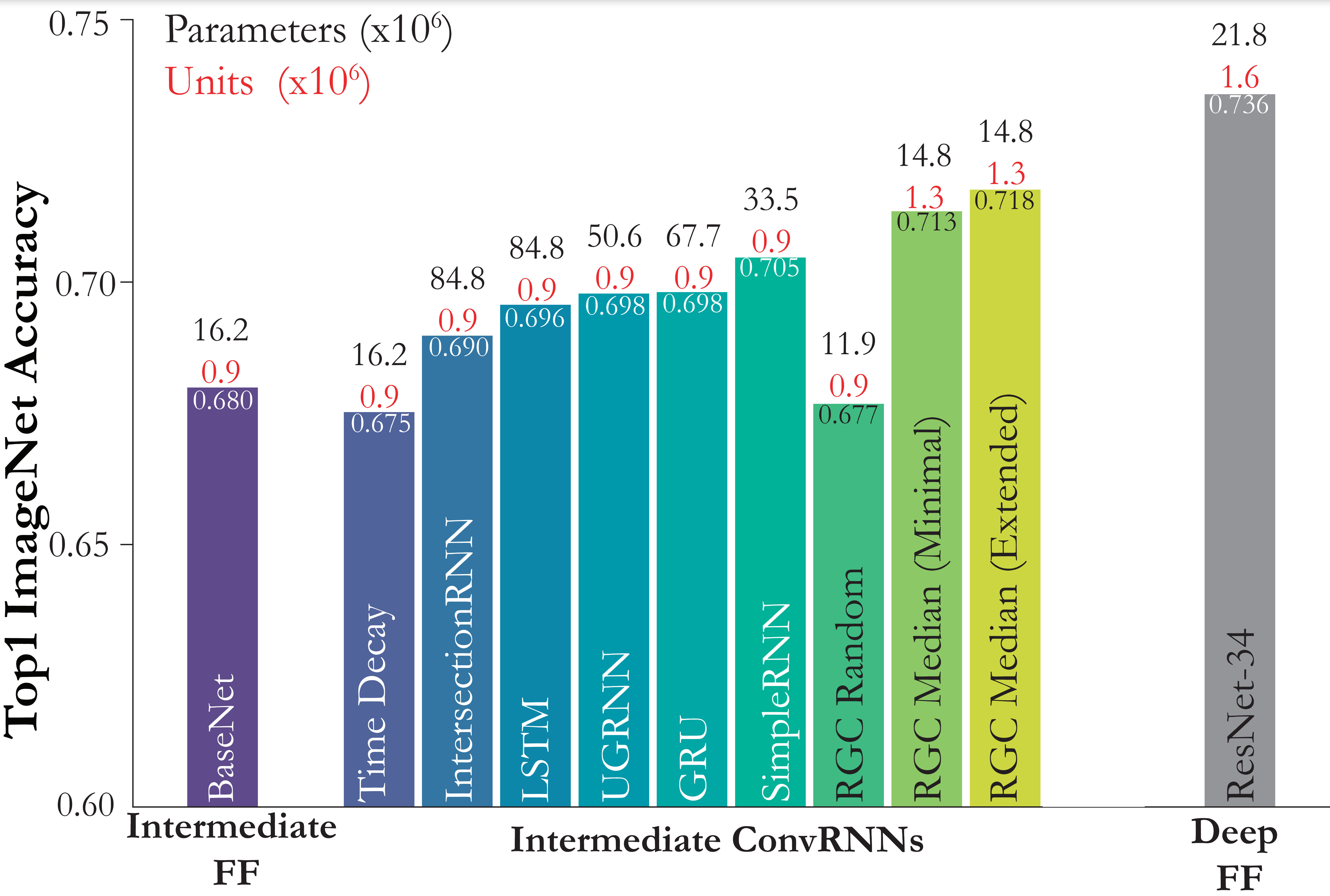


Emergent Local and Global Connectivity Patterns

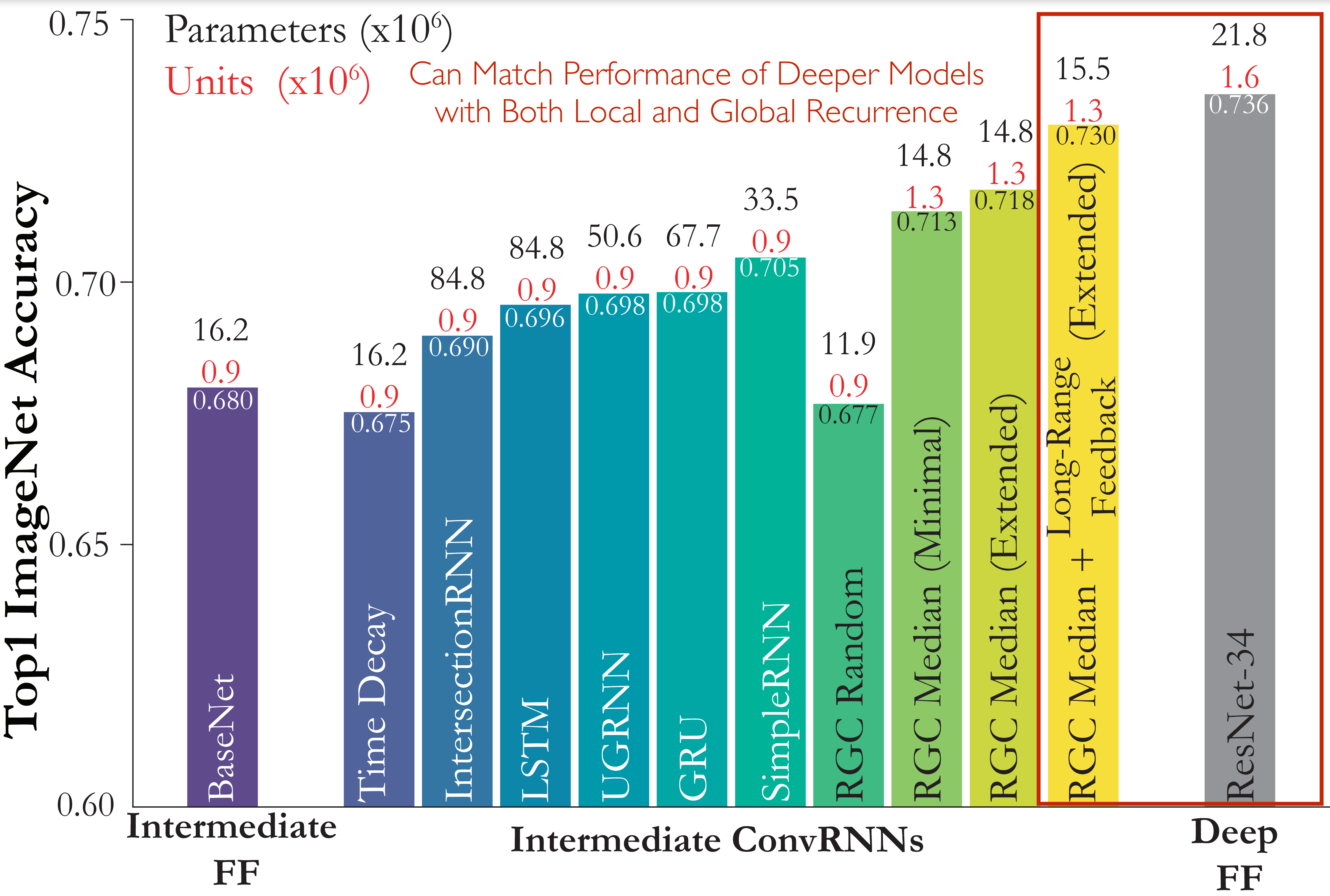


Conservation on parameter count as a byproduct of evolutionary optimization

Global Feedback Connections Matter



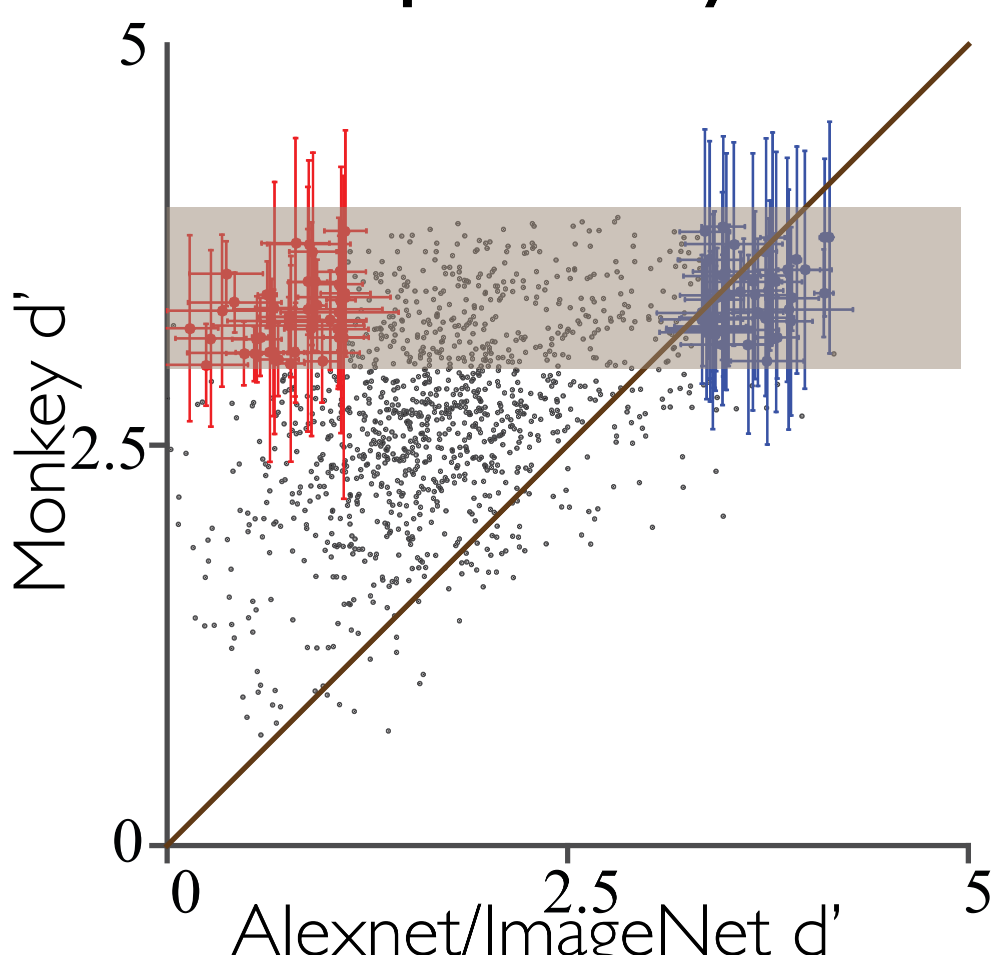
Global Feedback Connections Matter

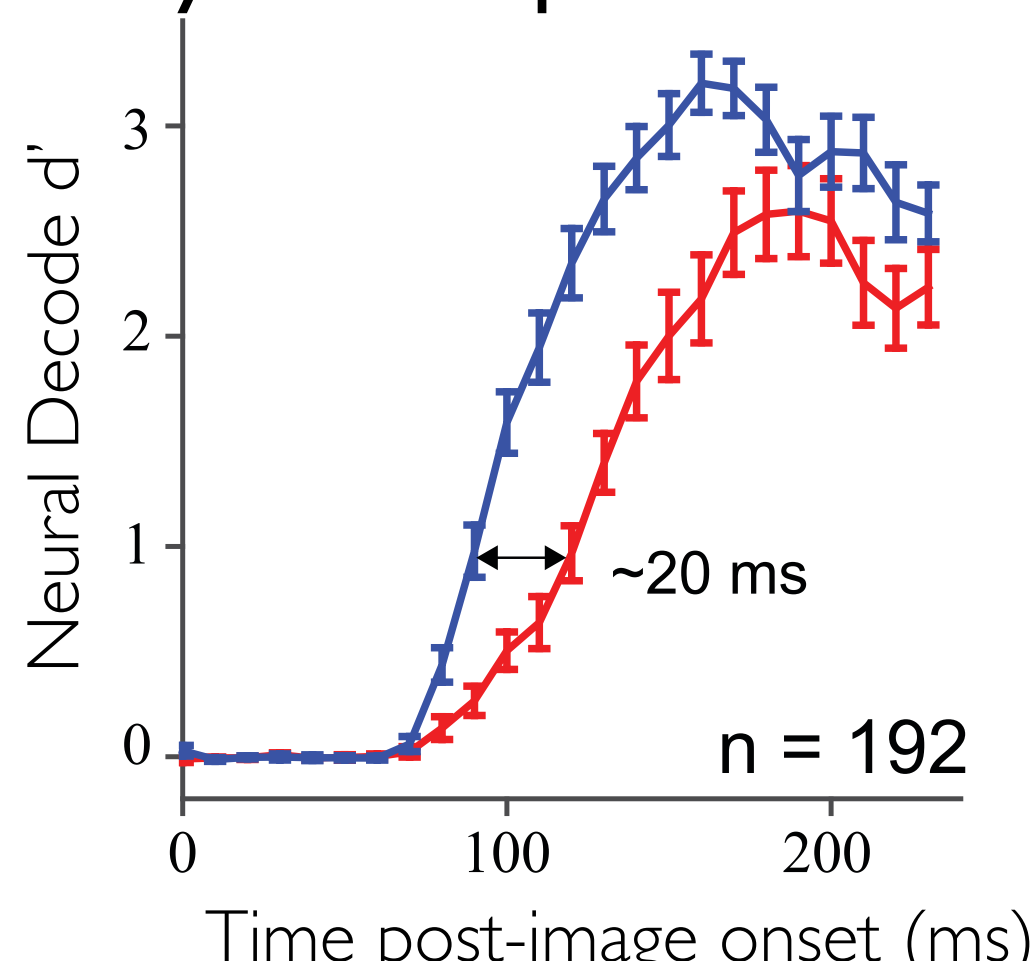


Comparing to Primate Object Solution Times (OSTs)

CNN-not-solved images *are* solved by the ventral visual stream

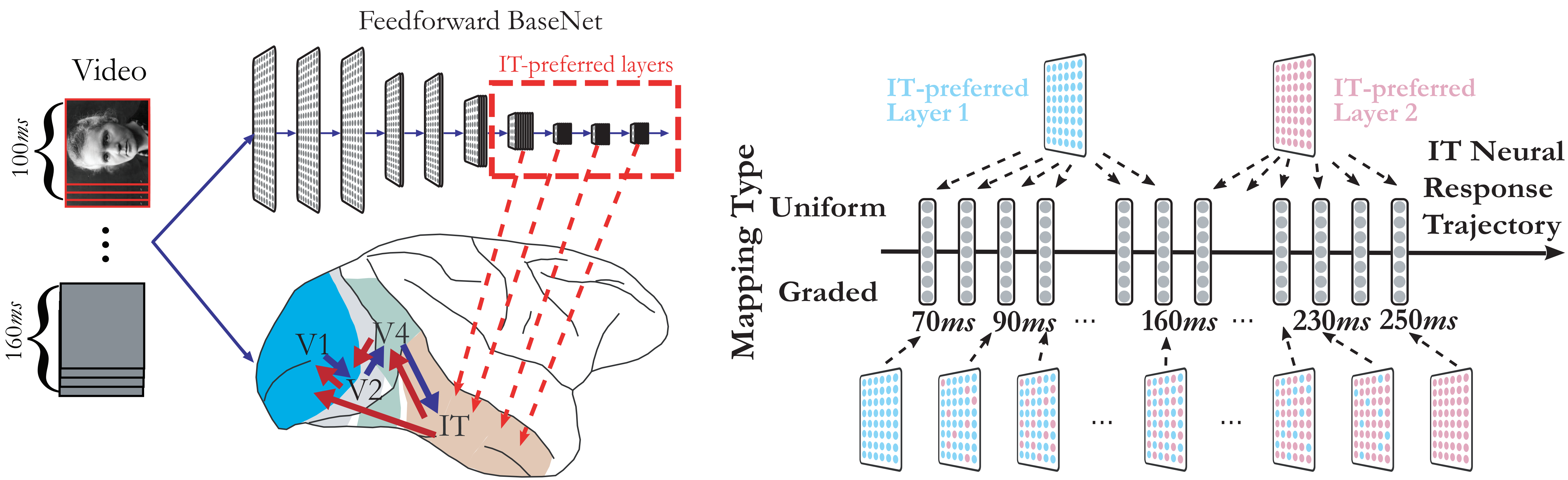
Can we explain *why* we see these dynamical patterns emerge in IT?



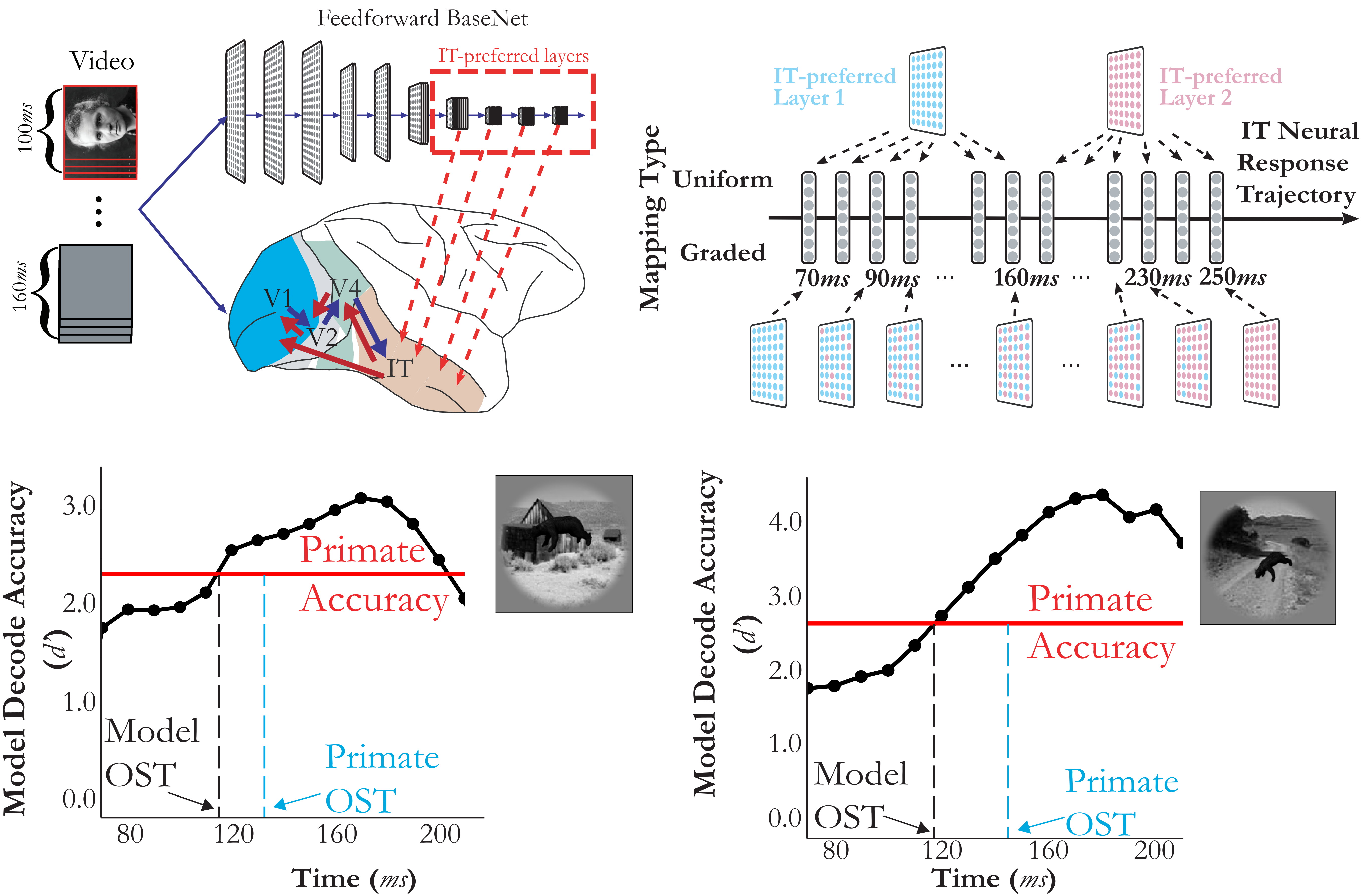


Kar et. al. (2019)

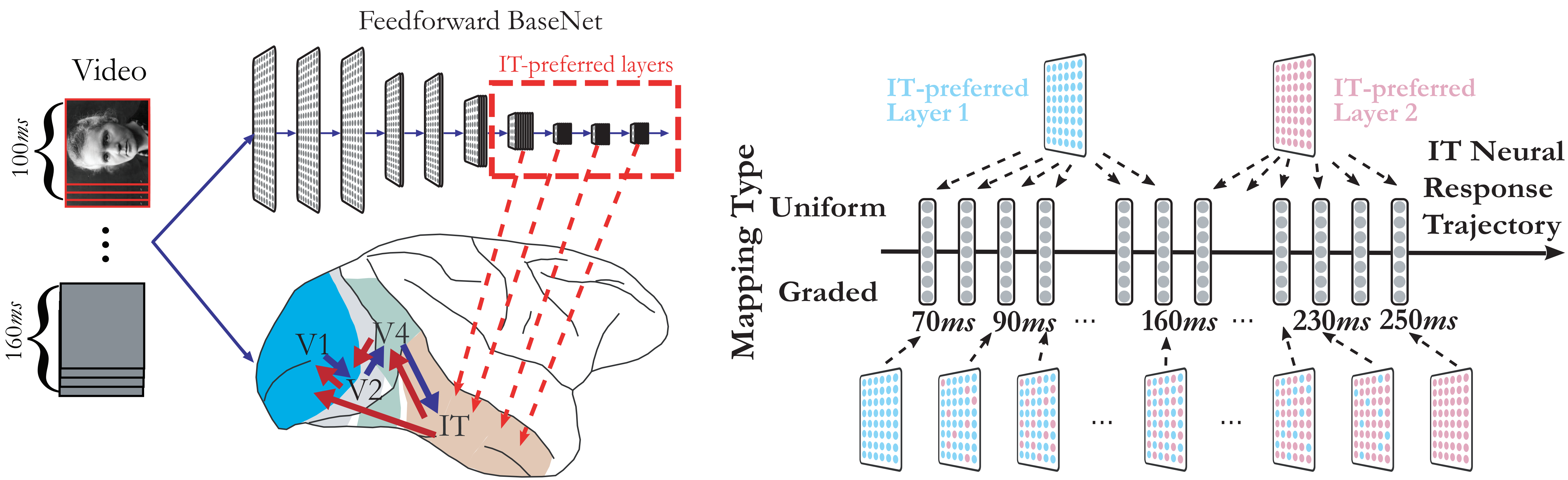
Comparing to Primate Object Solution Times (OSTs)



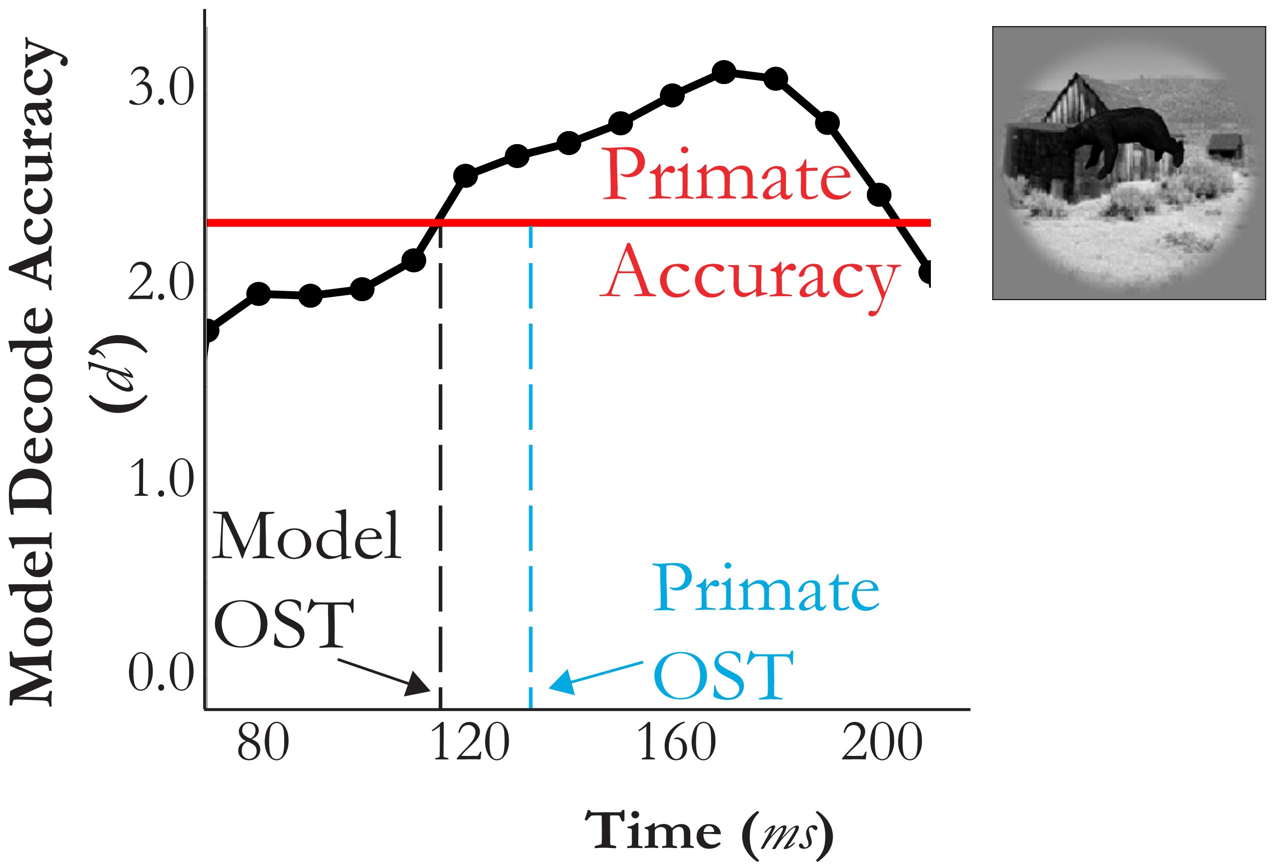
Comparing to Primate Object Solution Times (OSTs)

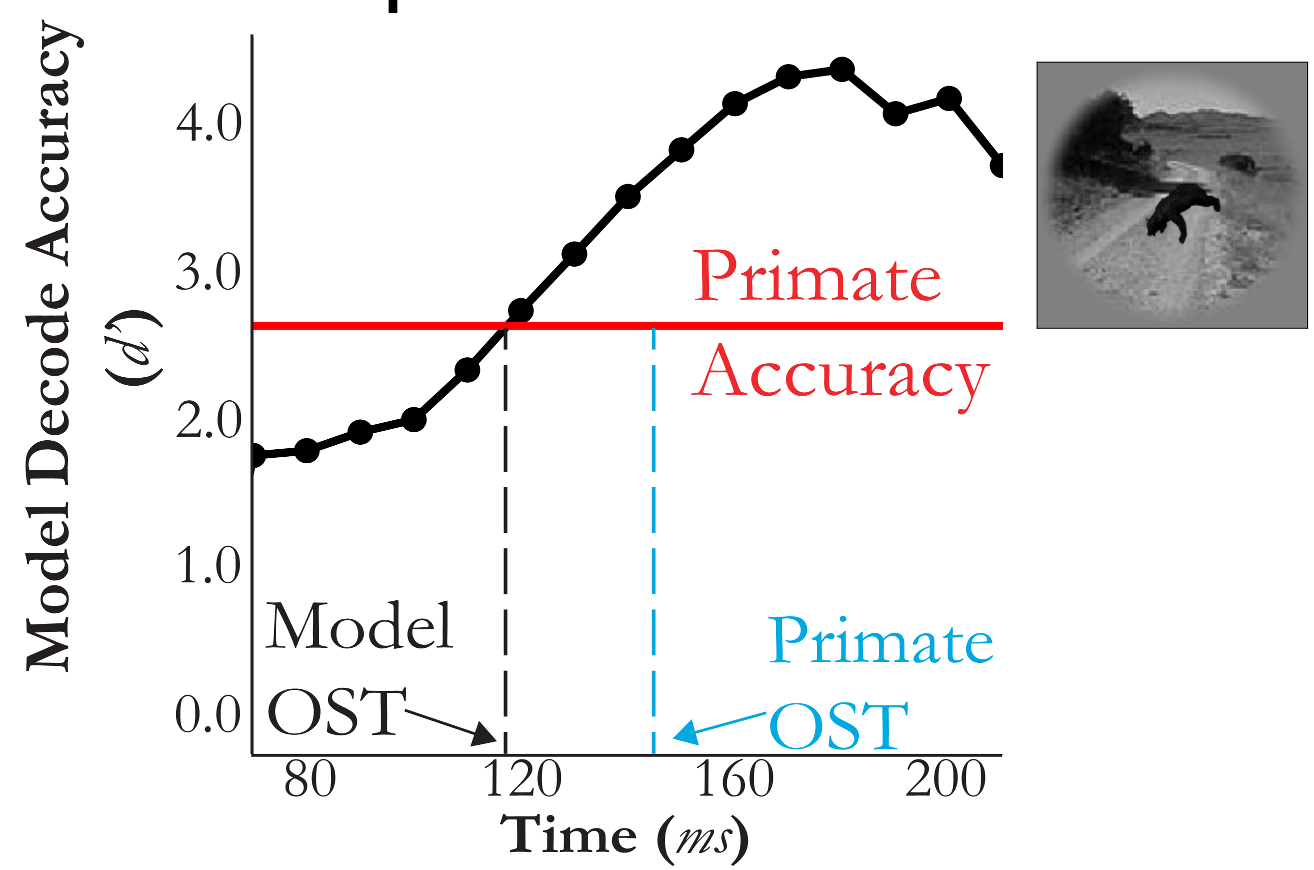


Comparing to Primate Object Solution Times (OSTs)

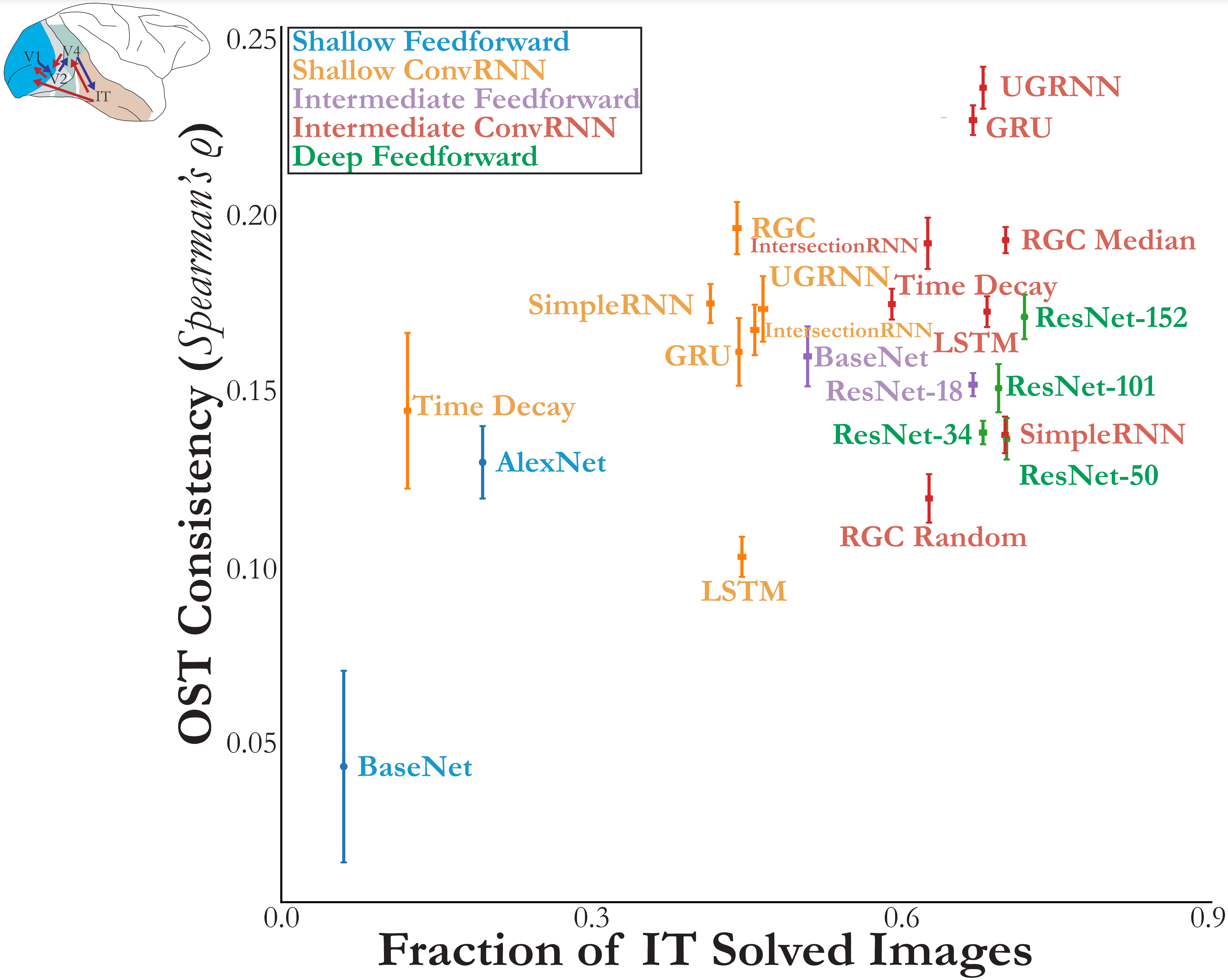


Both ConvRNNs and CNNs can be compared on this metric!

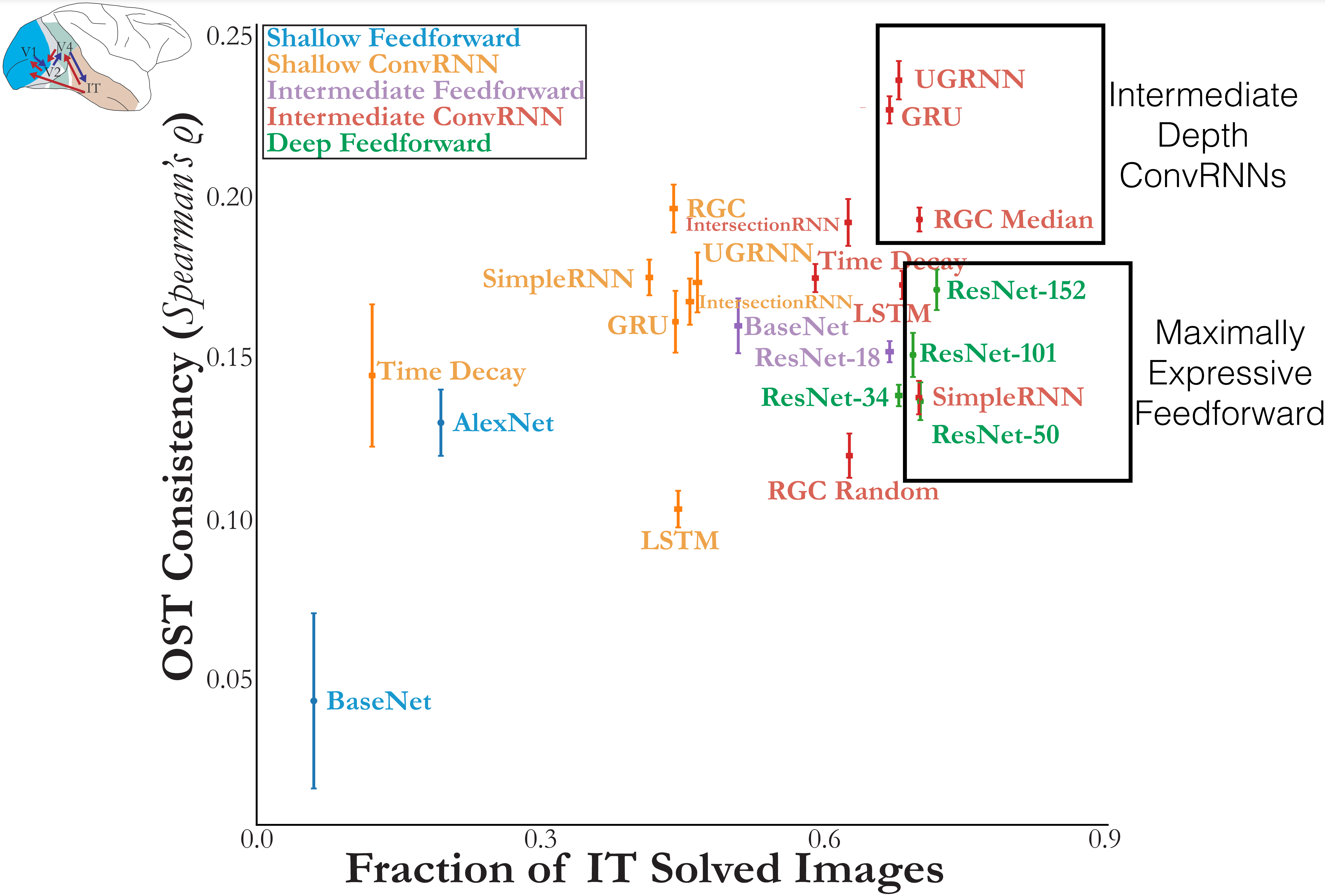




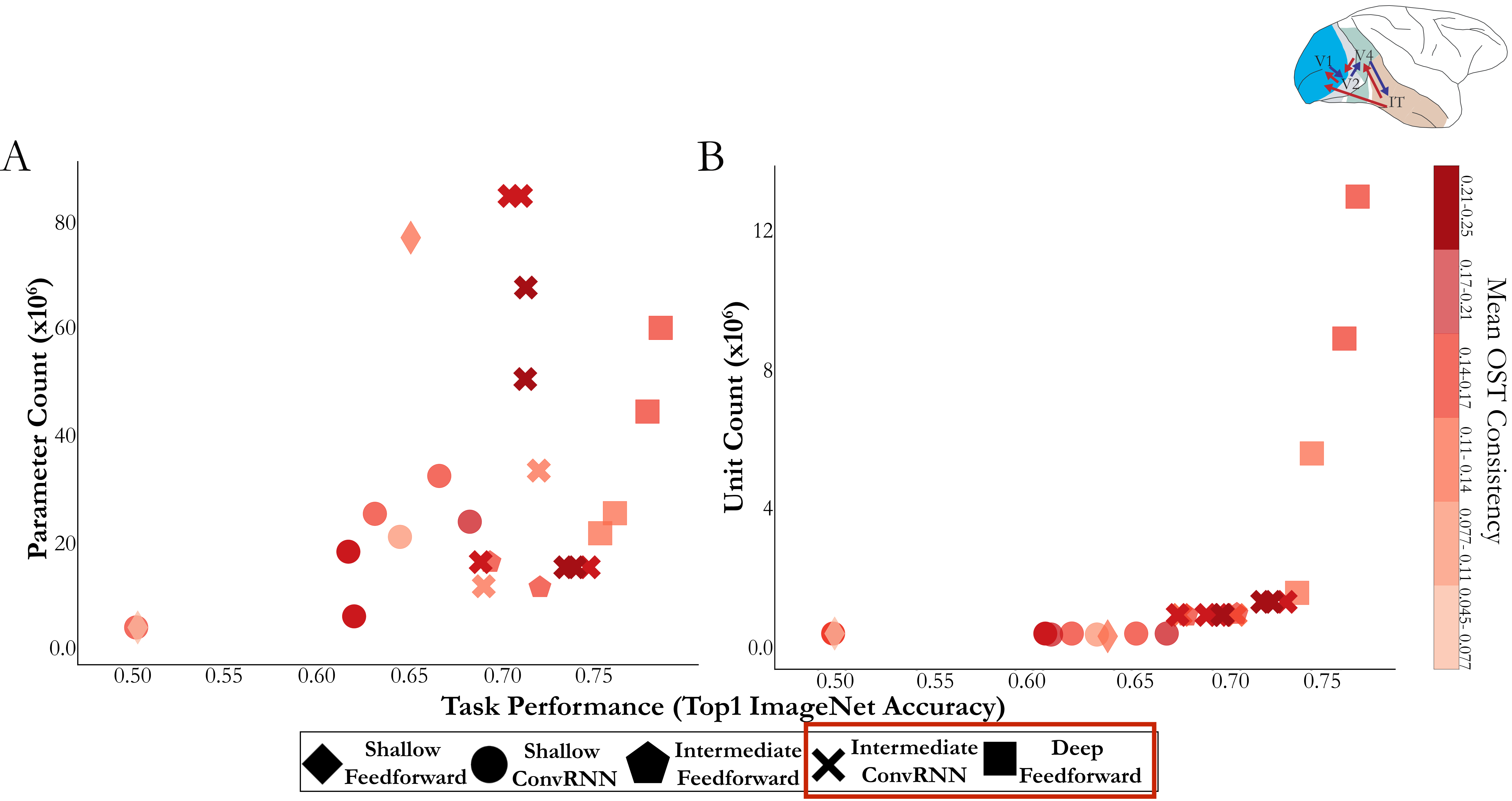
Intermediate ConvRNNs best match OSTs



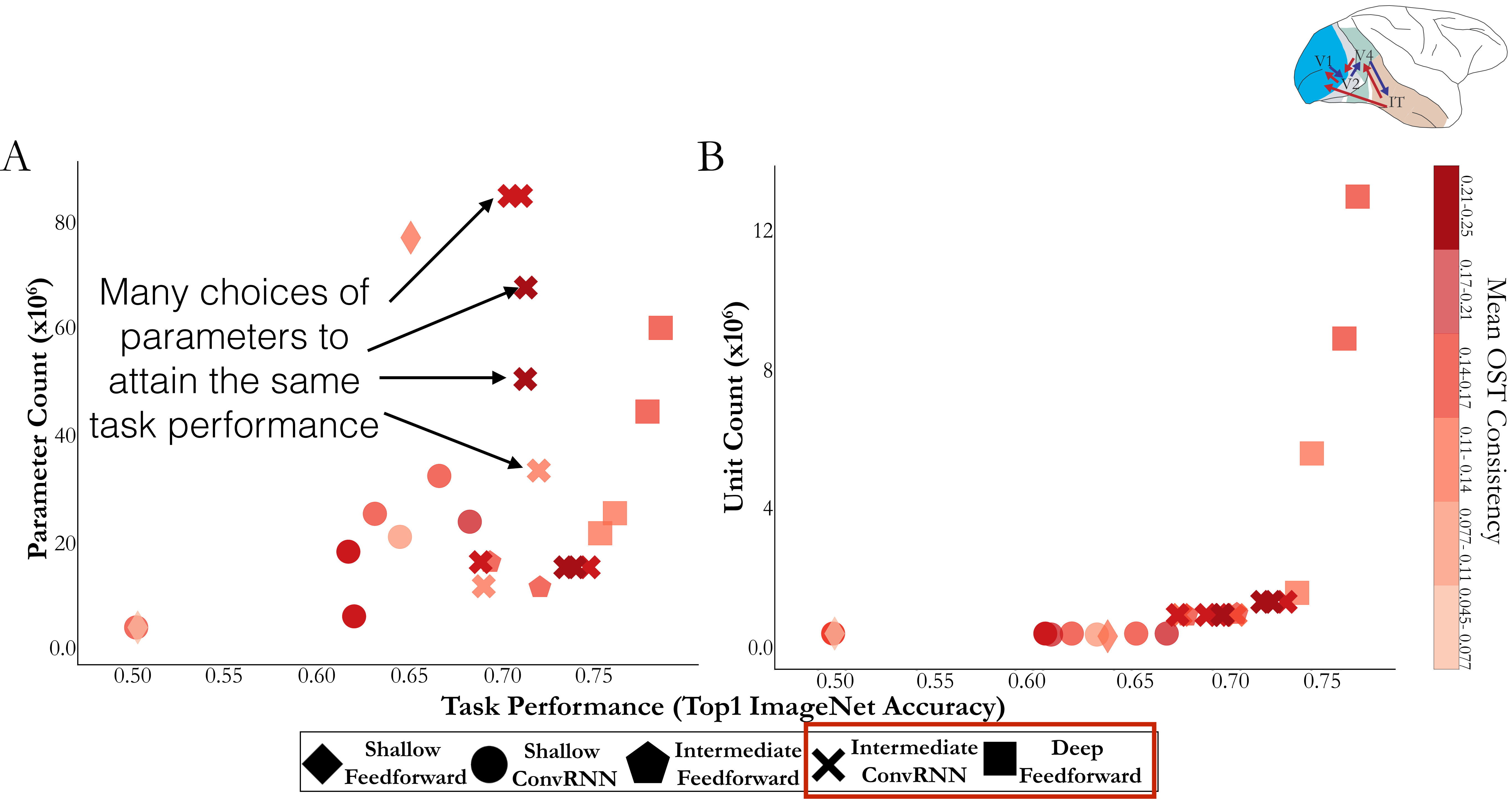
Intermediate ConvRNNs best match OSTs



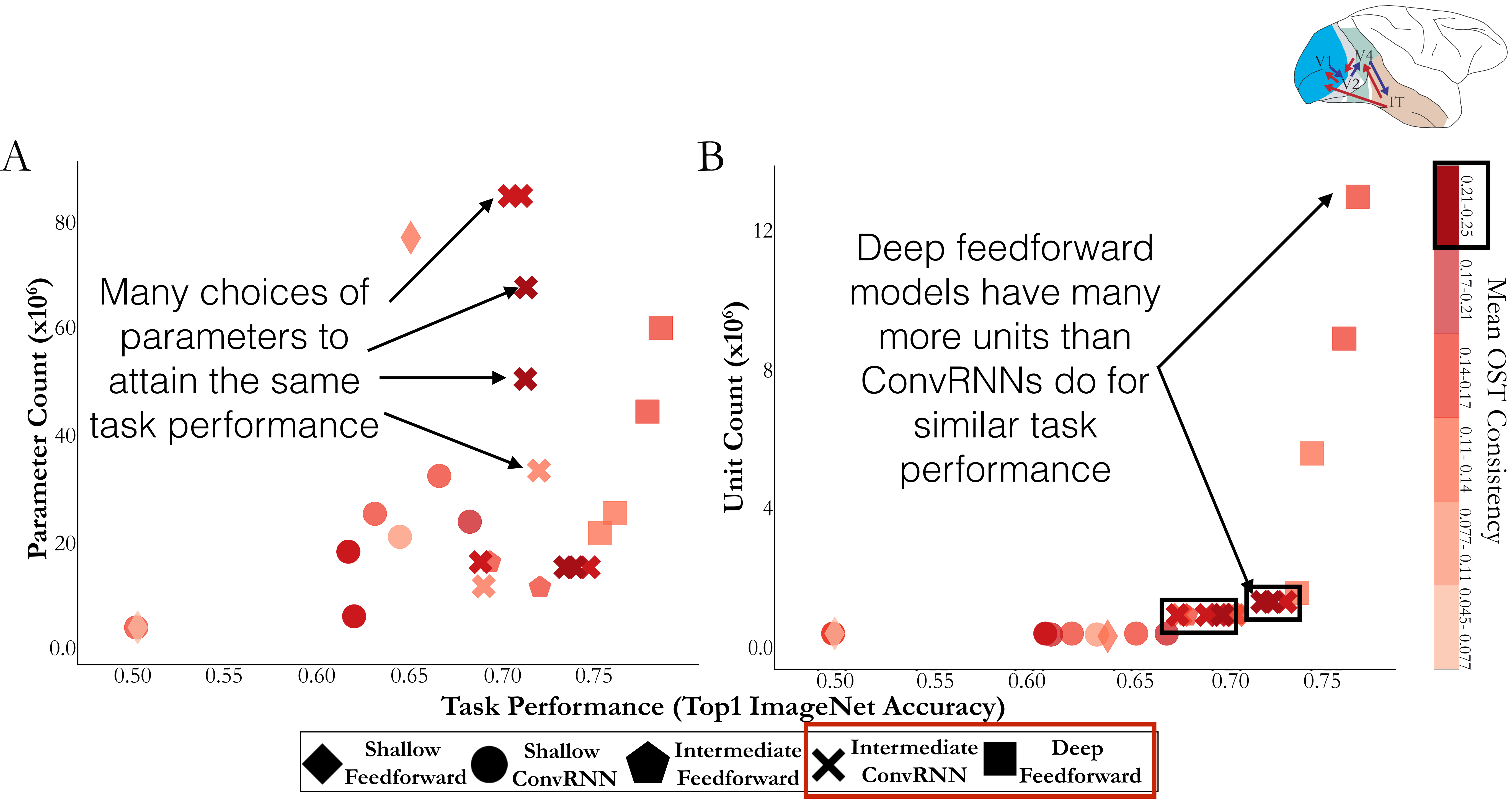
Conservation on network size + performance best matches OST



Conservation on network size + performance best matches OST



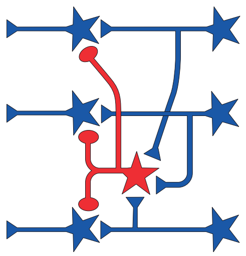
Conservation on network size + performance best matches OST



Takeaways

A = *architecture class*

1. *"Circuit"*

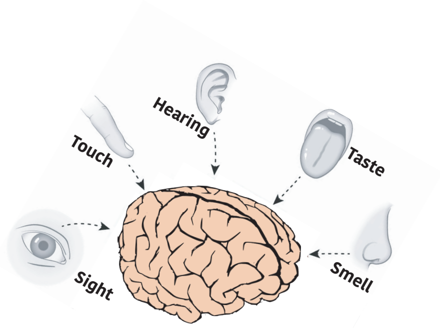


3. *"Ecological niche/behavior"*

T = *task loss*

Neurobiological Puzzle:

What is the role of recurrence in the primate ventral stream?



2. *"Environment"*

D = *data stream*

Takeaways

A = *architecture class*

1. *"Circuit"*

ConvRNNs

~~CNNs~~

T = *task loss*

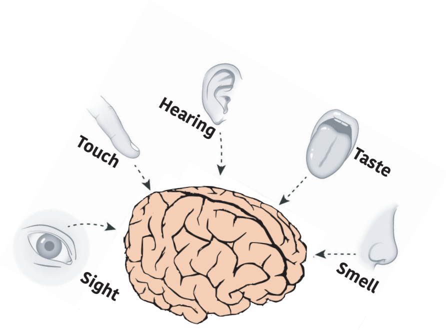
3. *"Ecological niche/behavior"*

<u>Neurobiological Puzzle:</u>

What is the role of recurrence in the primate ventral stream?

<u>Resolution:</u>

Enables high performance by trading off space with time, *in particular* space ~ # of neurons (<u>not</u> # of synapses).



2. *"Environment"*

D = *data stream*

Takeaways

A = *architecture class*

1. *"Circuit"*

ConvRNNs

~~CNNs~~

T = *task loss*

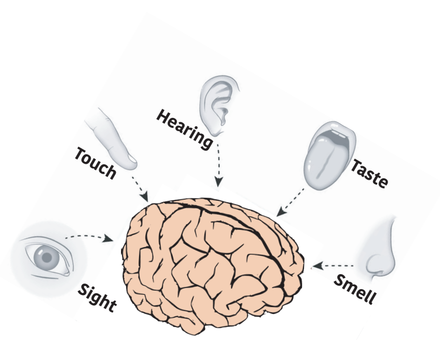
3. *"Ecological niche/behavior"*

<u>Neurobiological Puzzle:</u>

What is the role of recurrence in the primate ventral stream?

<u>*Partial* Resolution:</u>

Enables high performance by trading off space with time,
in particular space ~ # of neurons (<u>not</u> # of synapses).



2. *"Environment"*

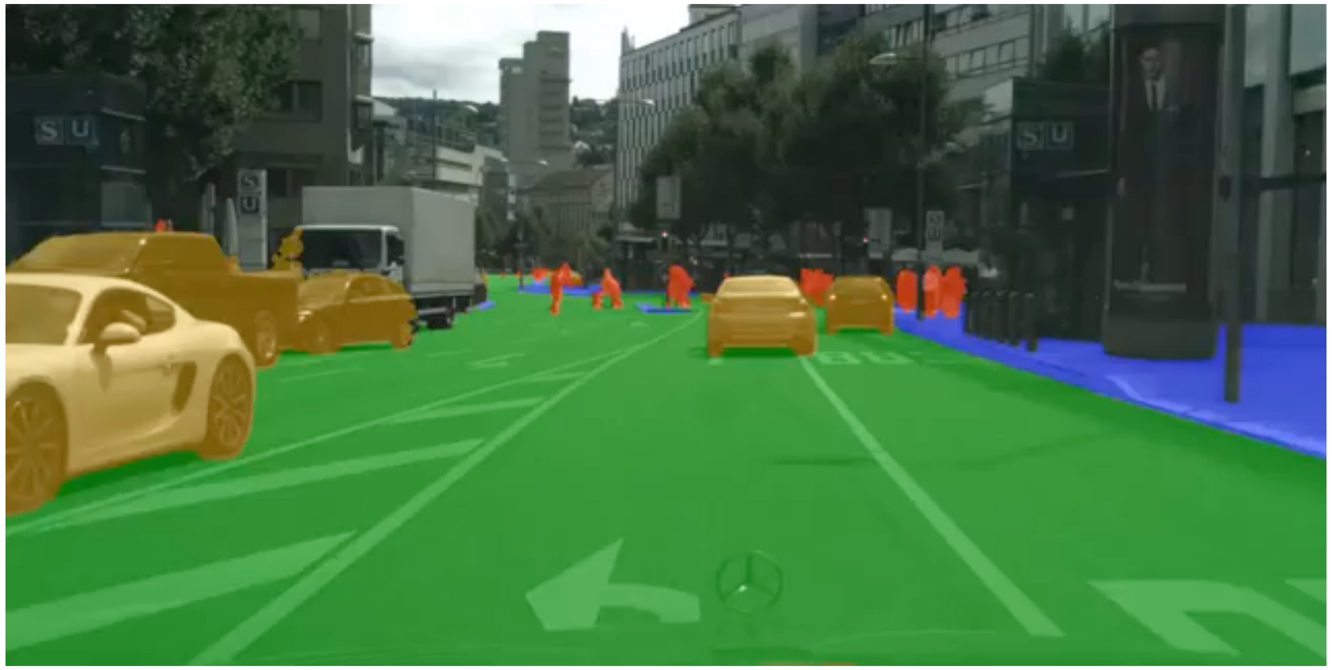
D = *data stream*

Outline

- Recurrent Connections in the Primate Ventral Stream
- Goal-Driven Models of Mouse Visual Cortex
- Heterogeneity in Rodent Medial Entorhinal Cortex
- Future Directions

From Neurons to Behavior

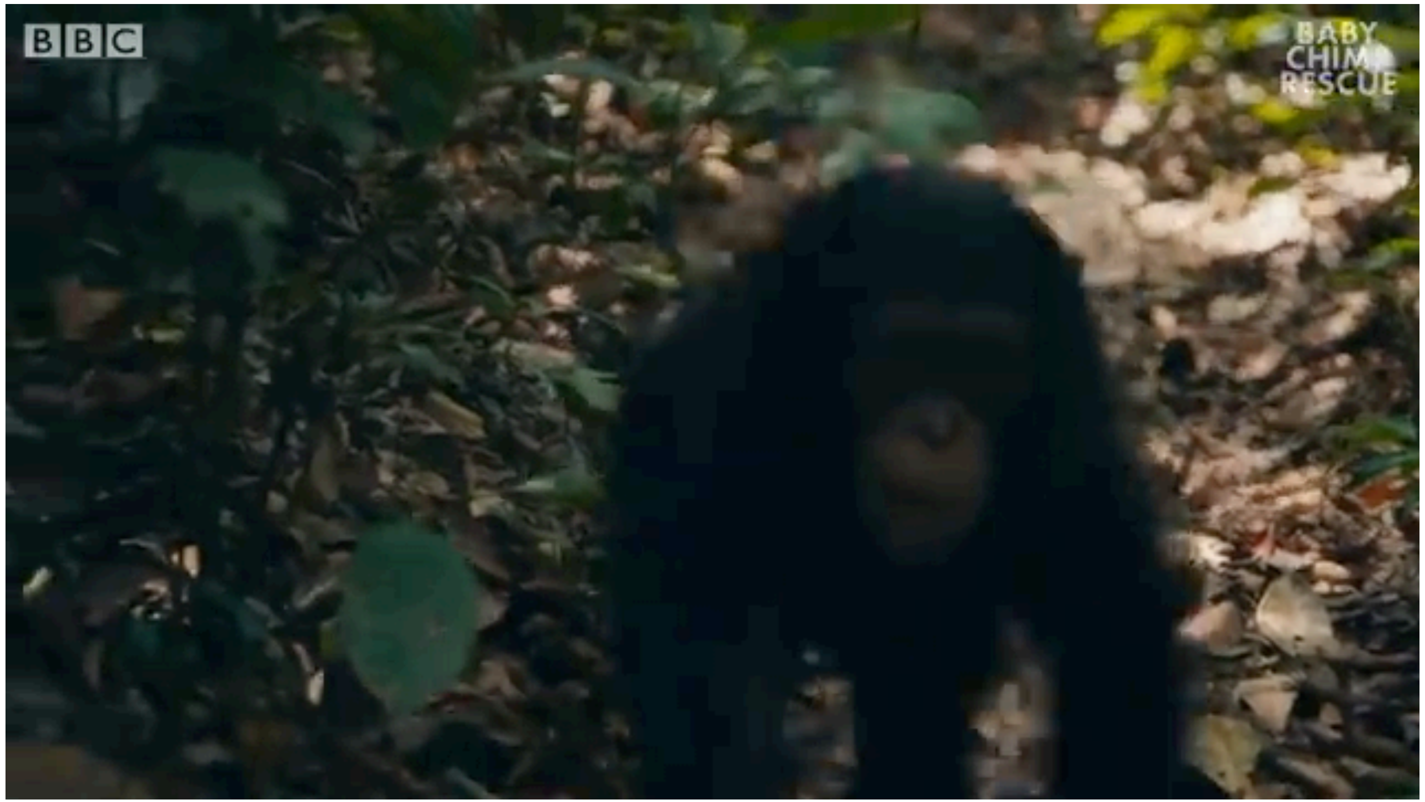
Scene Understanding



Multi-Step Planning

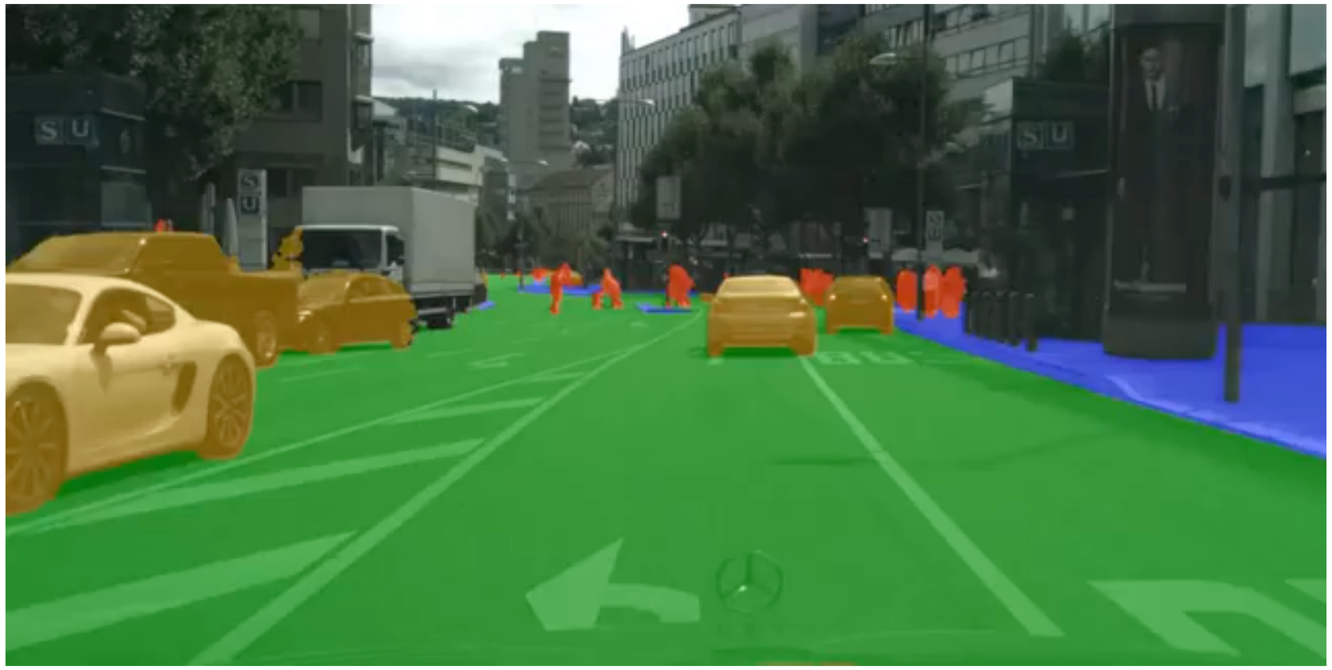
Navigation

Flexible Embodiment



From Neurons to Behavior

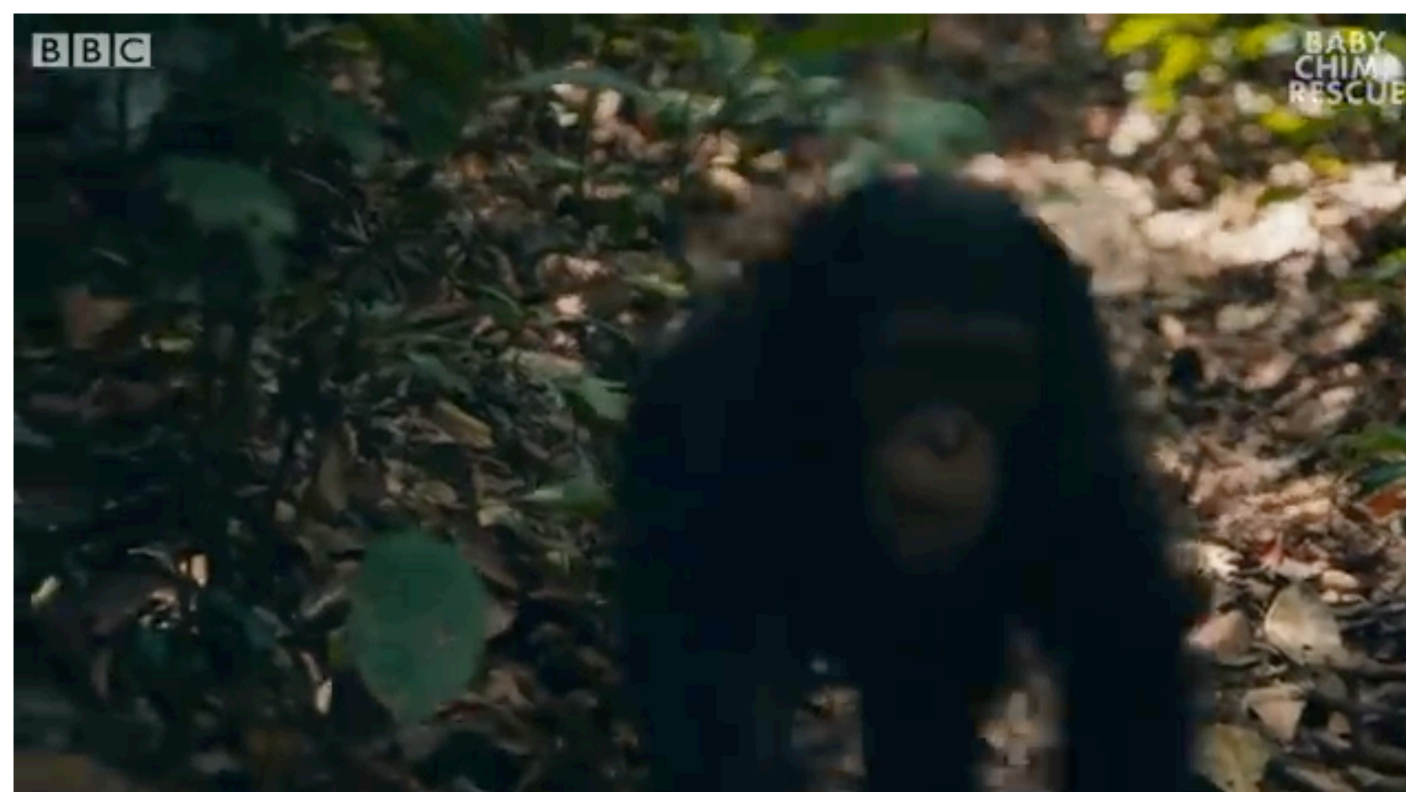
Scene Understanding



Multi-Step Planning

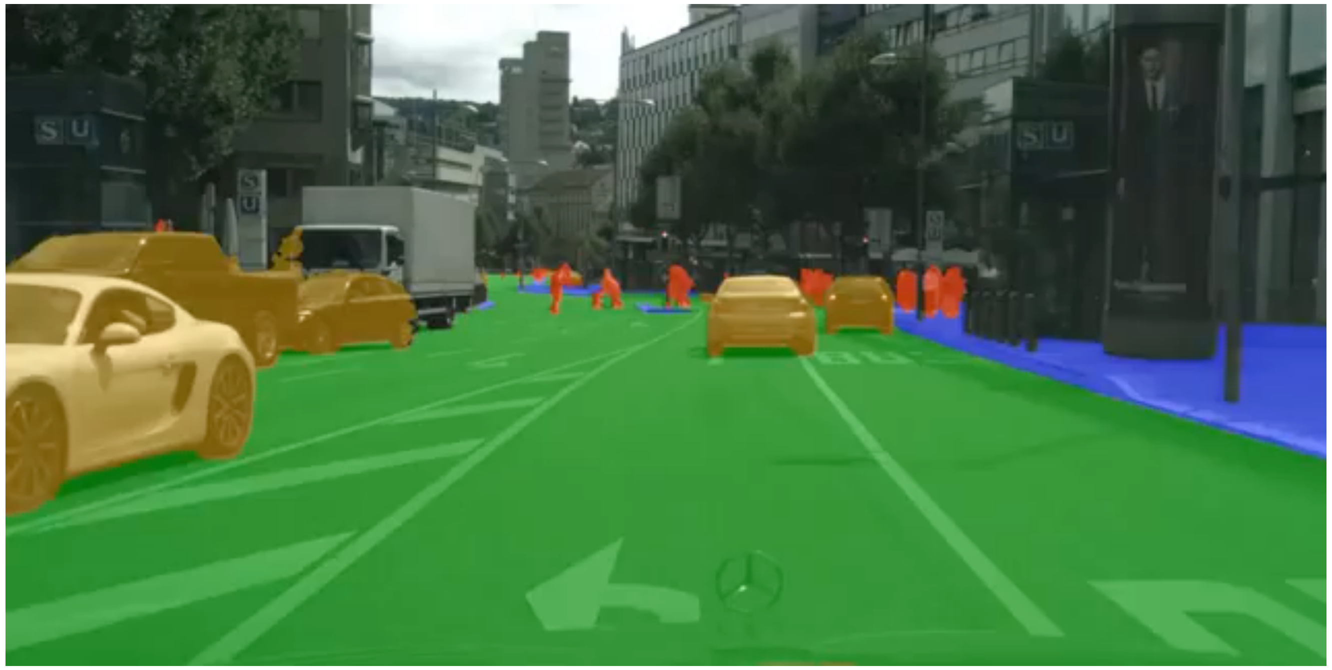
Navigation

Flexible Embodiment



From Neurons to Behavior

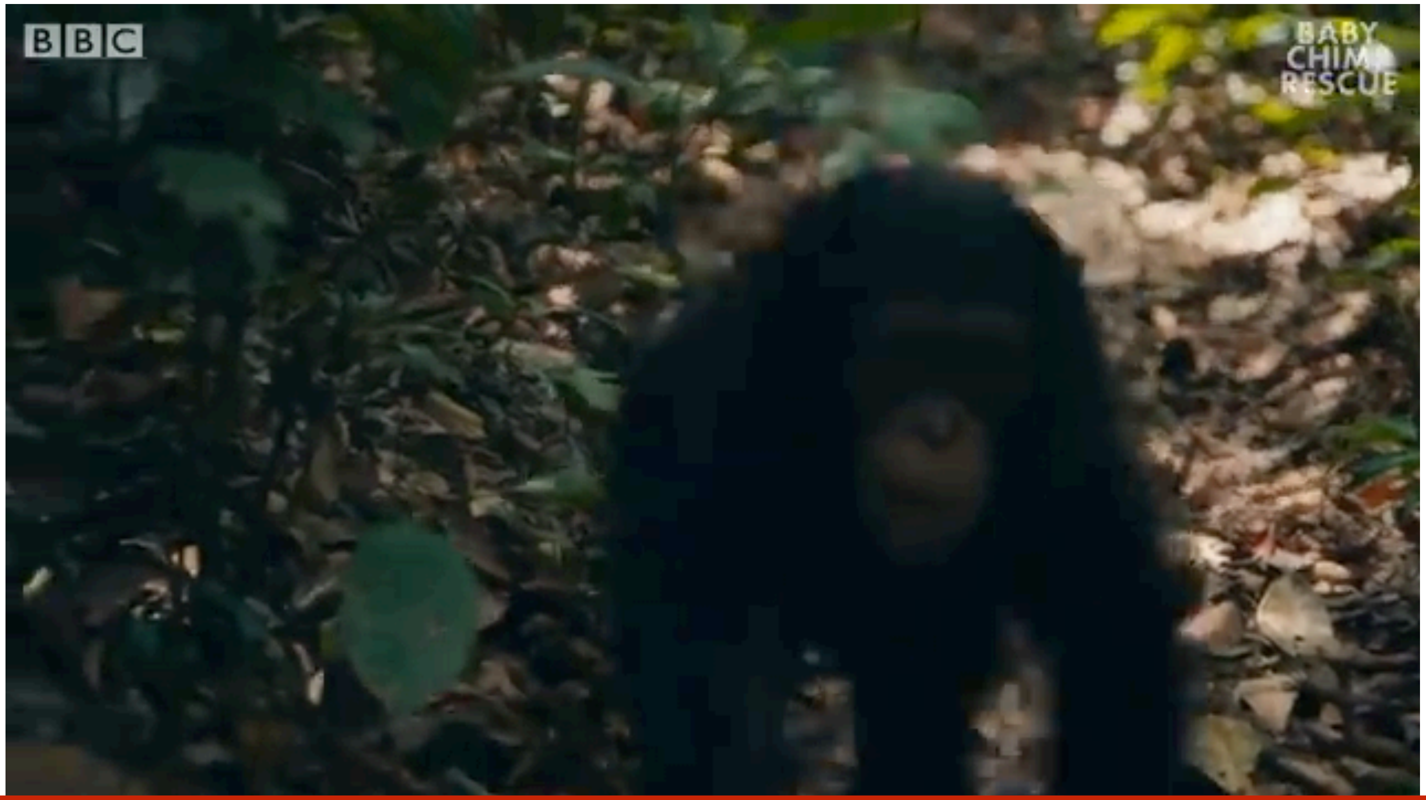
Scene Understanding



Multi-Step Planning

Navigation

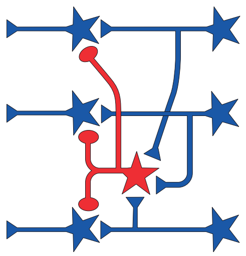
Flexible Embodiment



Goal-Driven Models of Mouse Visual Cortex

A *= architecture class*

1. *"Circuit"*



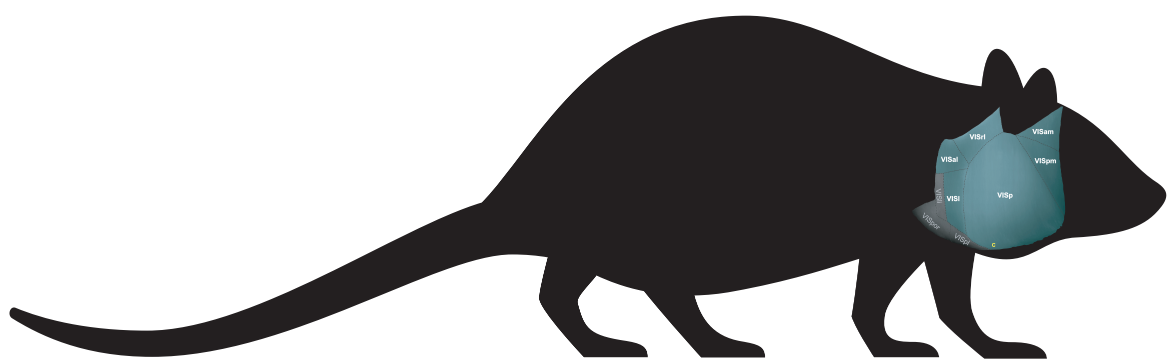
T *= task loss*

3. *"Ecological niche/behavior"*

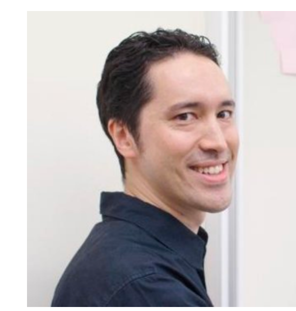
A. Nayebi*, N. Kong* *et al.*

Mouse visual cortex as a limited resource system that self-learns an ecologically-general representation. *bioRxiv (2021)*

Nathan C.L. Kong*



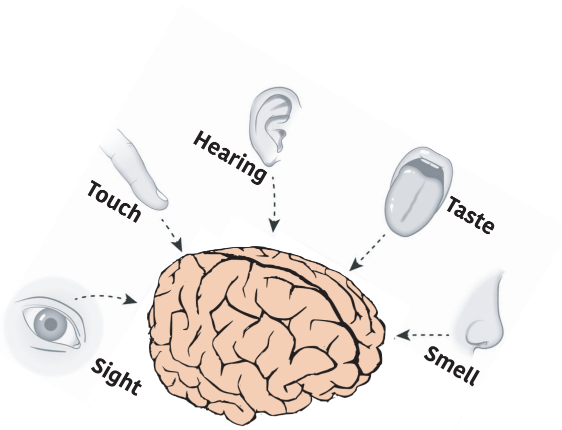
Chengxu Zhuang



Justin L. Gardner

Anthony M. Norcia

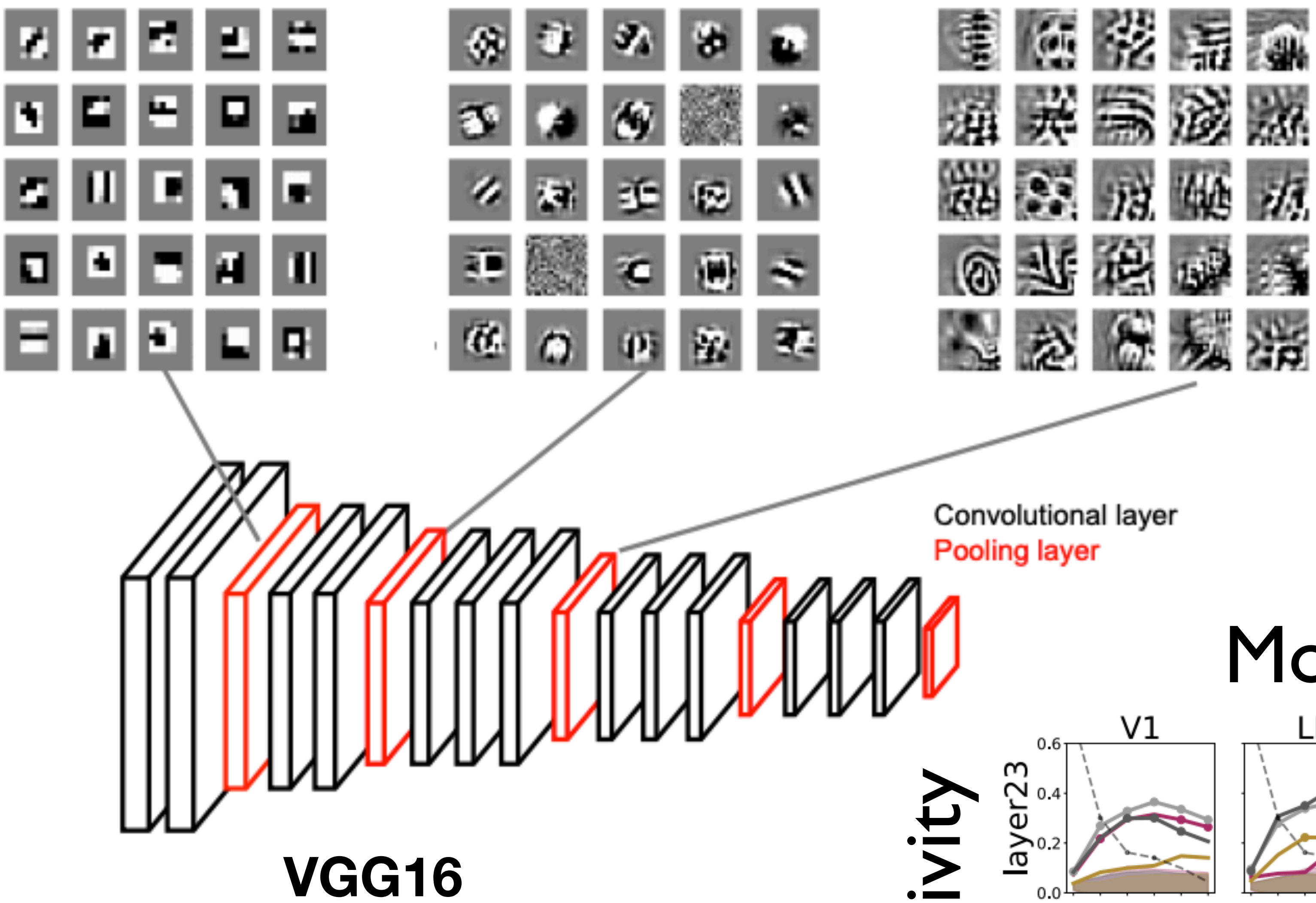
Daniel Yamins

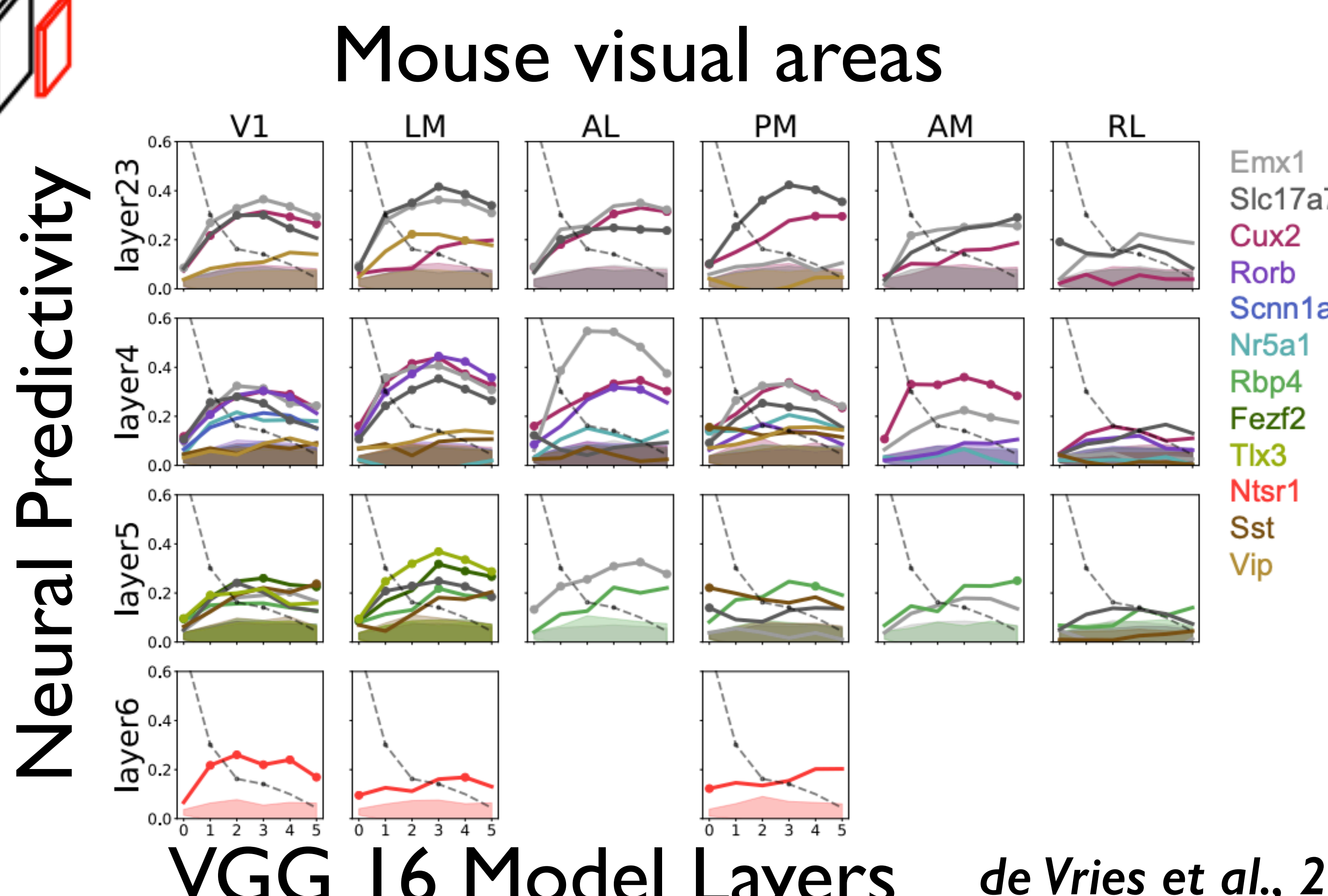


2. *"Environment"*

D *= data stream*

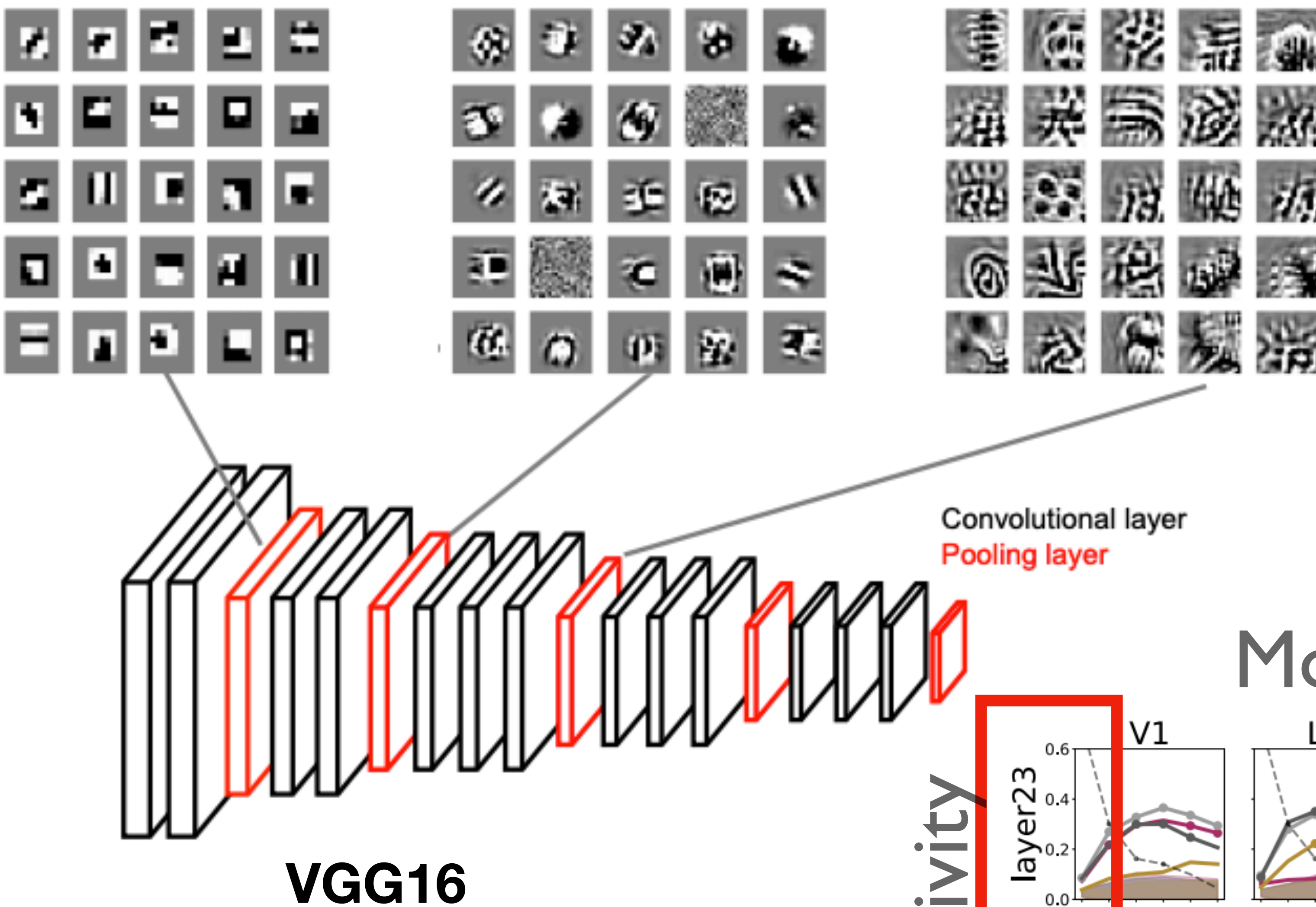
Initial deep neural network models of mouse visual cortex



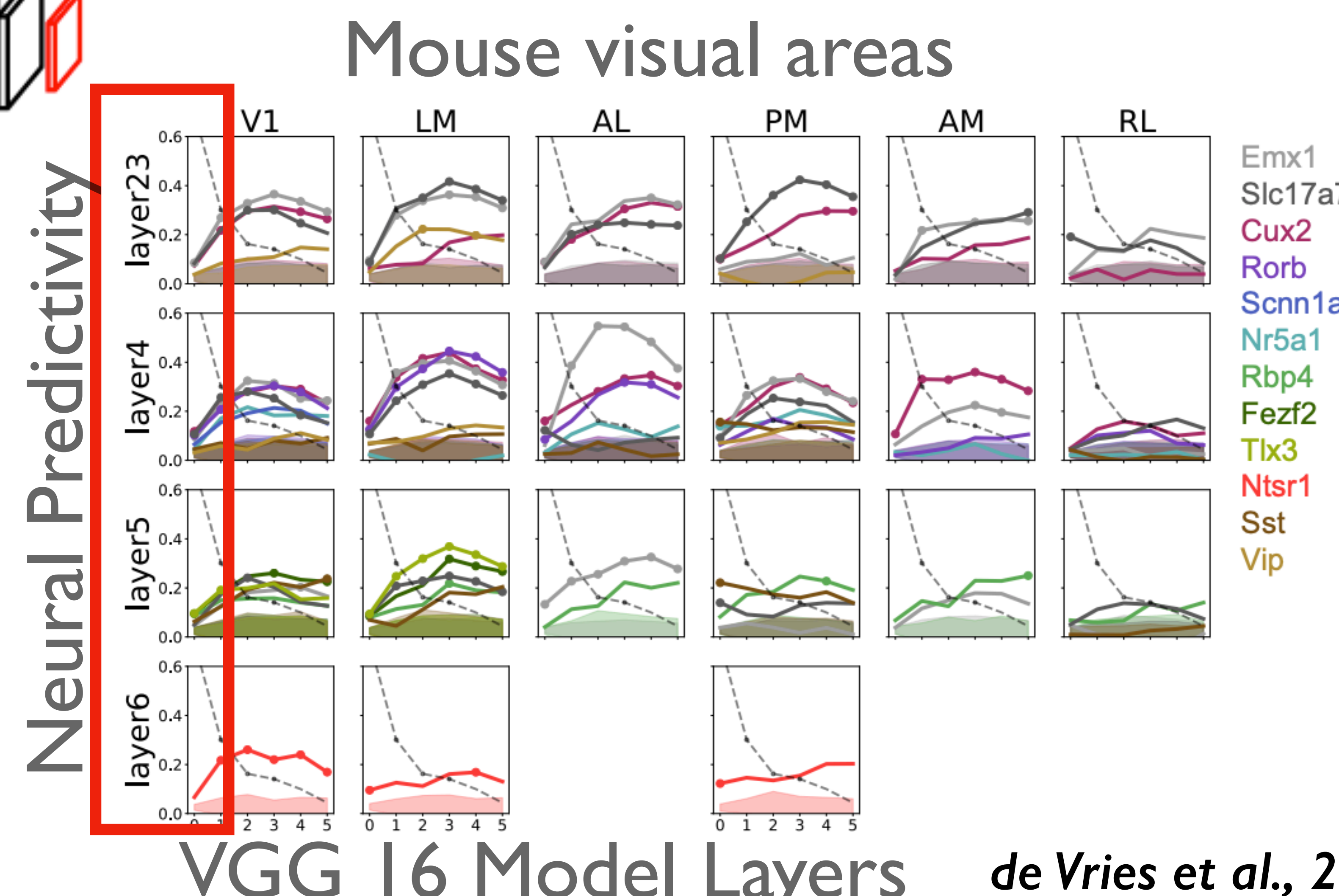


de Vries et al., 2020

Deep models are a poor match to responses

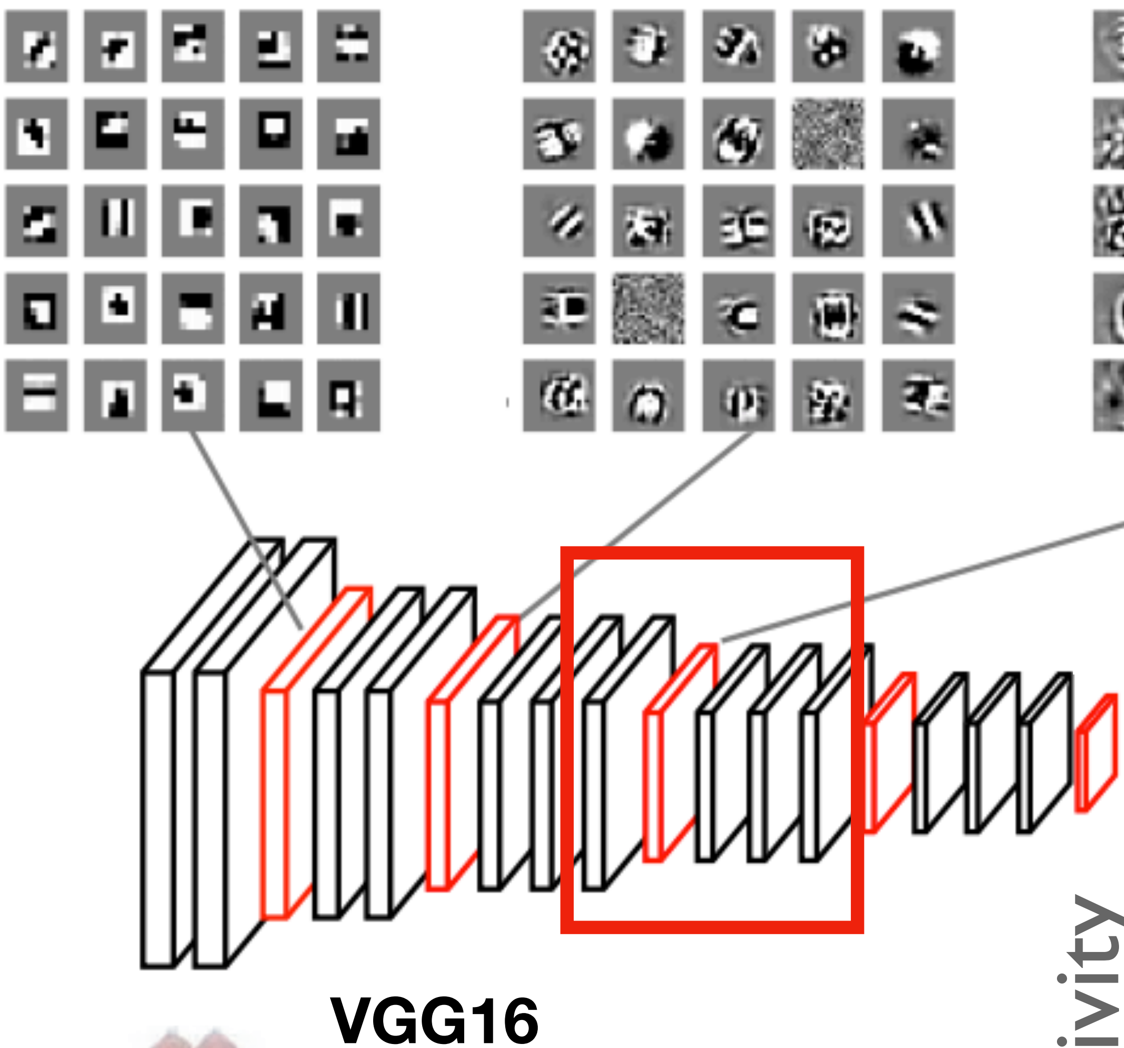


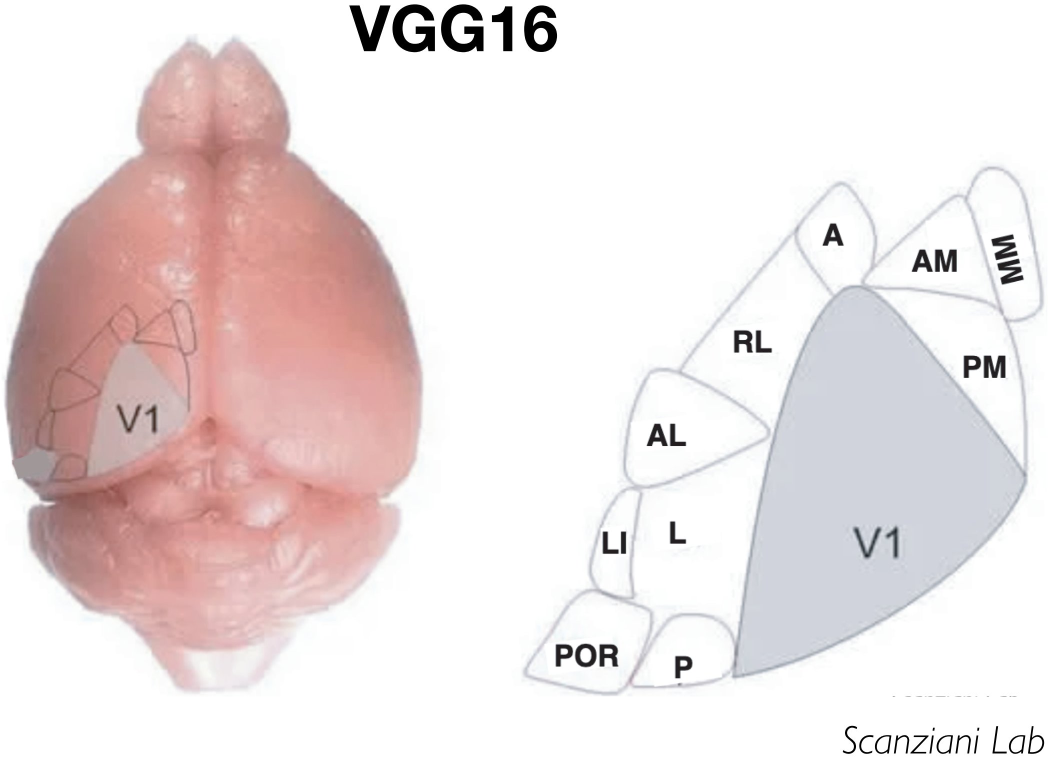
Low predictivity between model's layers and mouse visual areas!



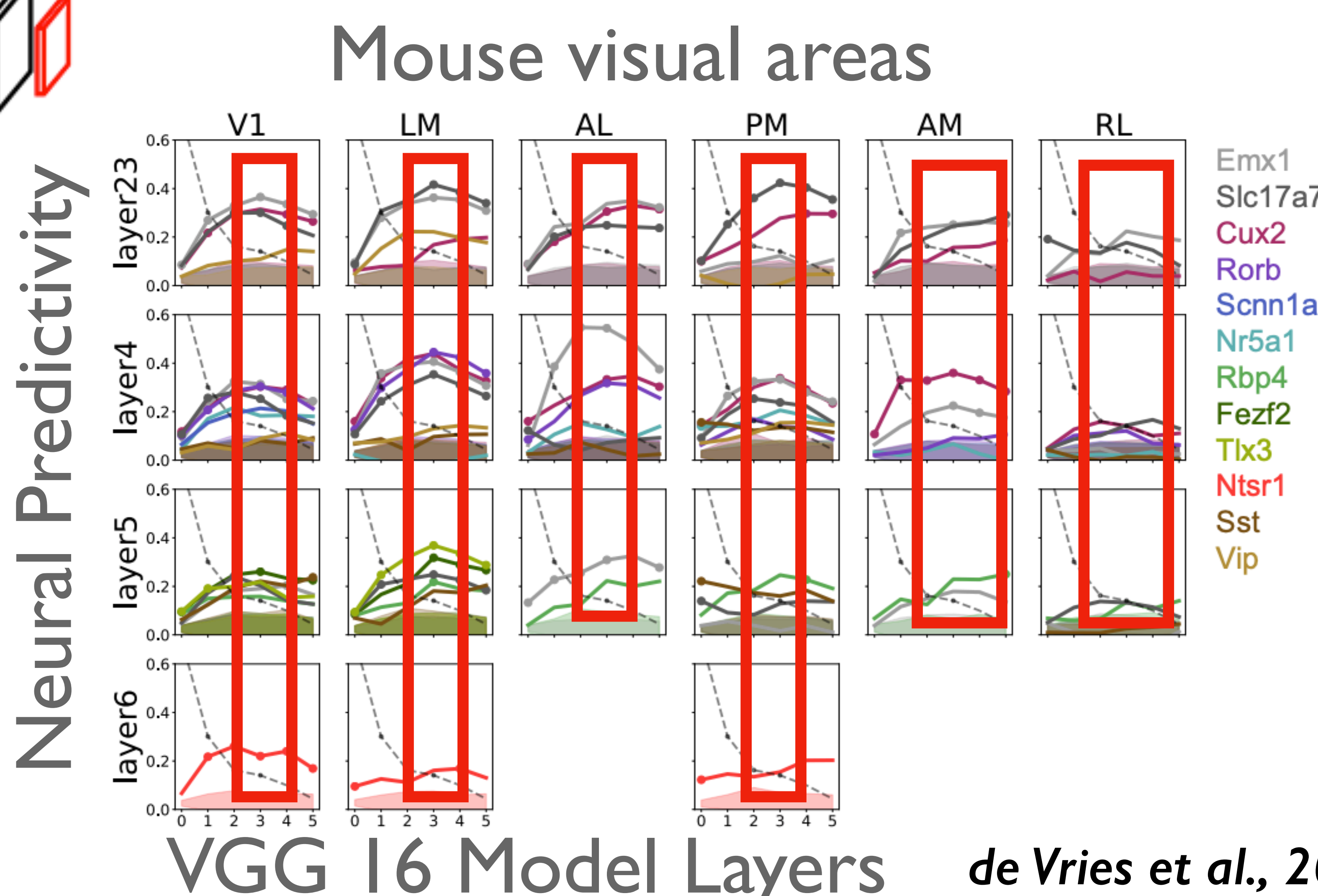
de Vries et al., 2020

Deep models suggest mouse visual cortex is representationally deep





Mid-levels best predict visual areas

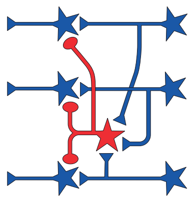


de Vries et al., 2020

Goal-Driven Models of Mouse Visual Cortex

A *= architecture class*

1. *"Circuit"*

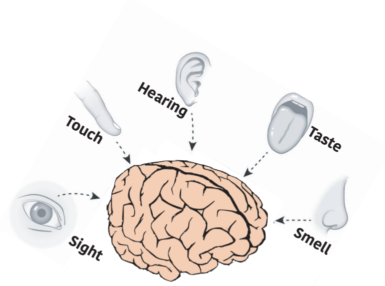


L *= loss function*

3. *"Ecological niche/behavior"*

<u>Neurobiological Puzzle:</u>

What are the core differences between mouse visual cortex and the primate ventral stream?



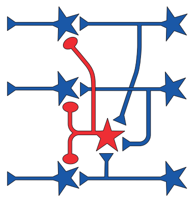
2. *"Environment"*

D *= data stream*

Putting it all together: Circuit, Inputs, Behavior

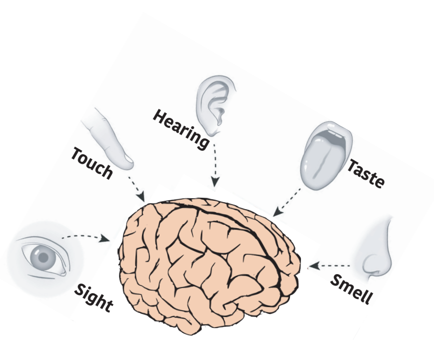
A = *architecture class*

1. *"Circuit"*



T = *task loss*

3. *"Ecological niche/behavior"*



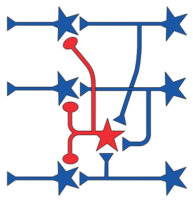
2. *"Environment"*

D = *data stream*

Putting it all together: Circuit, Inputs, Behavior

A *= architecture class*

1. *"Circuit"*

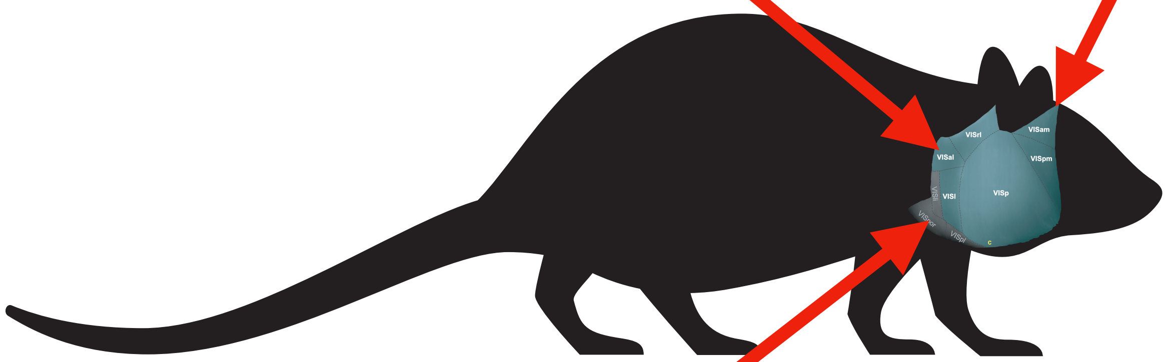


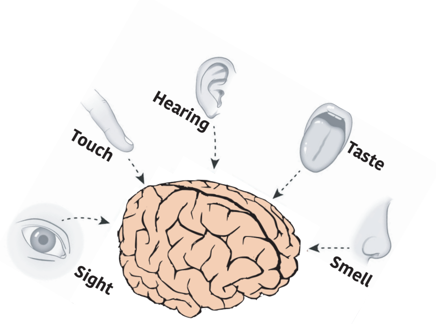
shallow

~~deep~~

T *= task loss*

3. *"Ecological niche/behavior"*





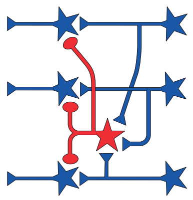
2. *"Environment"*

D *= data stream*

Putting it all together: Circuit, Inputs, Behavior

A = *architecture class*

1. *"Circuit"*



shallow

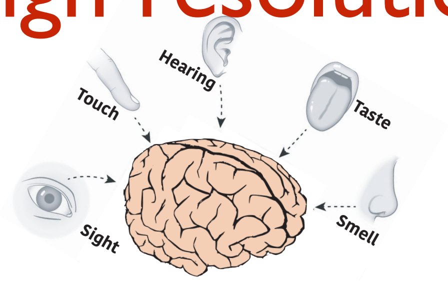
~~deep~~

T = *task loss*

3. *"Ecological niche/behavior"*

low resolution

~~high resolution~~



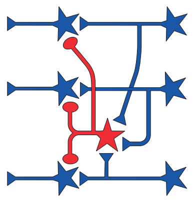
2. *"Environment"*

D = *data stream*

Putting it all together: Circuit, Inputs, Behavior

A = *architecture class*

1. *"Circuit"*



shallow

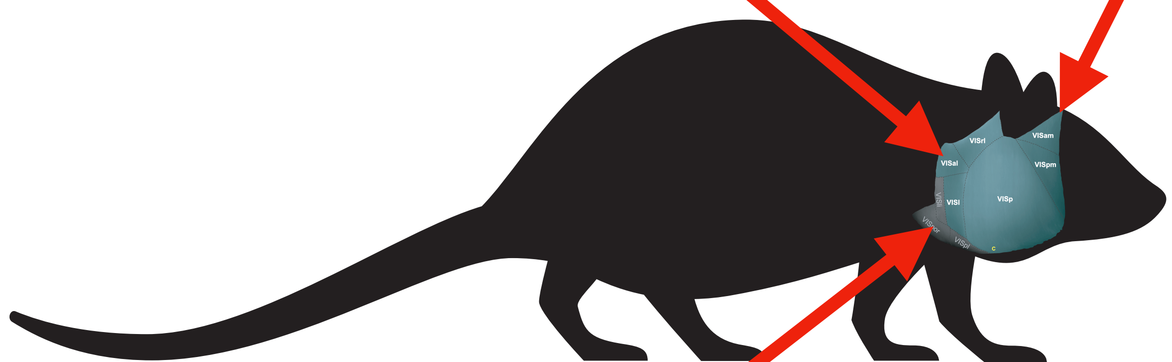
~~deep~~

T = *task loss*

3. *"Ecological niche/behavior"*

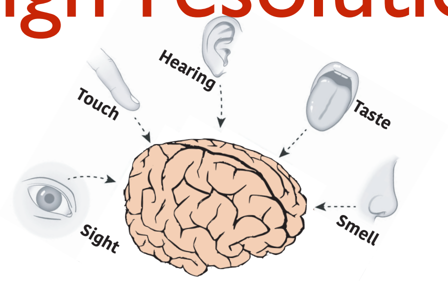
unsupervised

~~supervised~~



low resolution

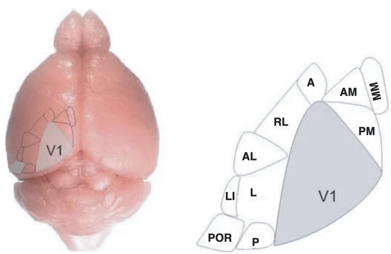
~~high resolution~~

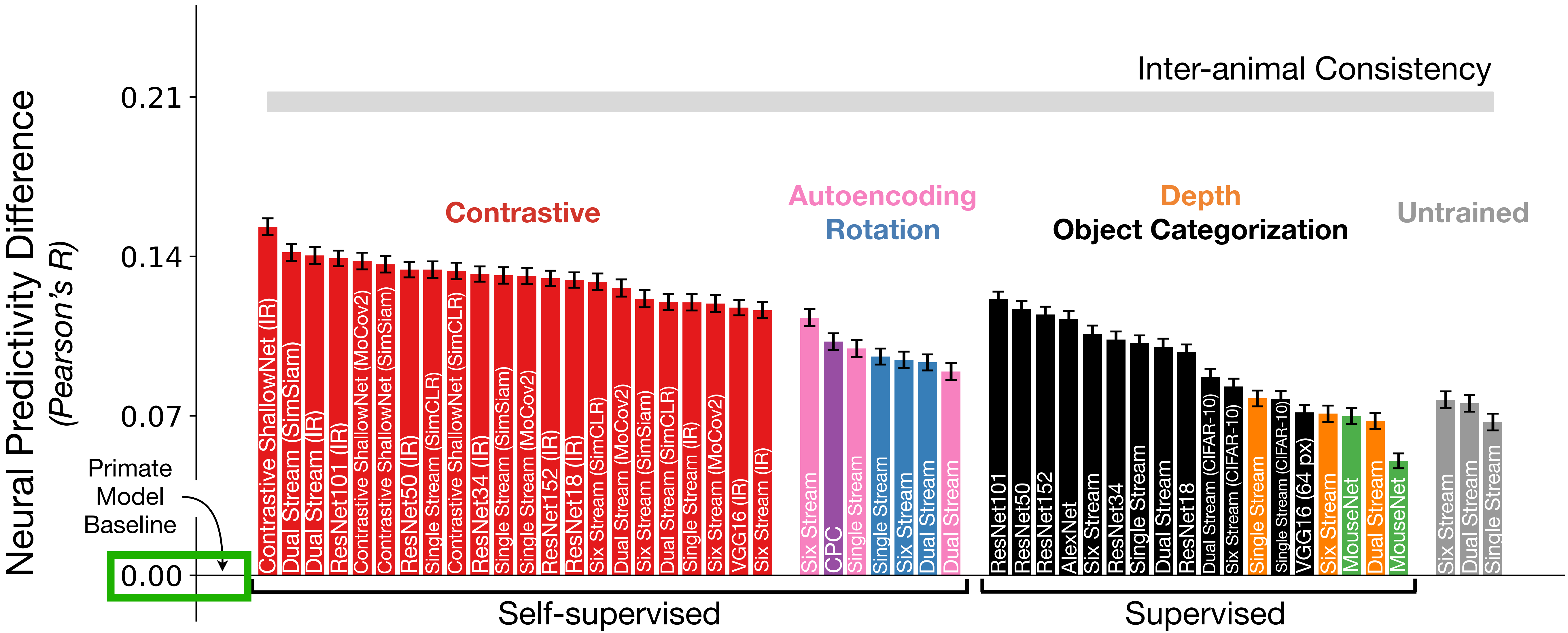


2. *"Environment"*

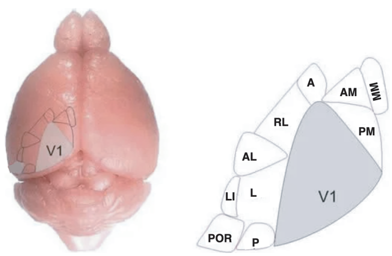
D = *data stream*

Substantially improving neural response predictivity of models of mouse visual cortex

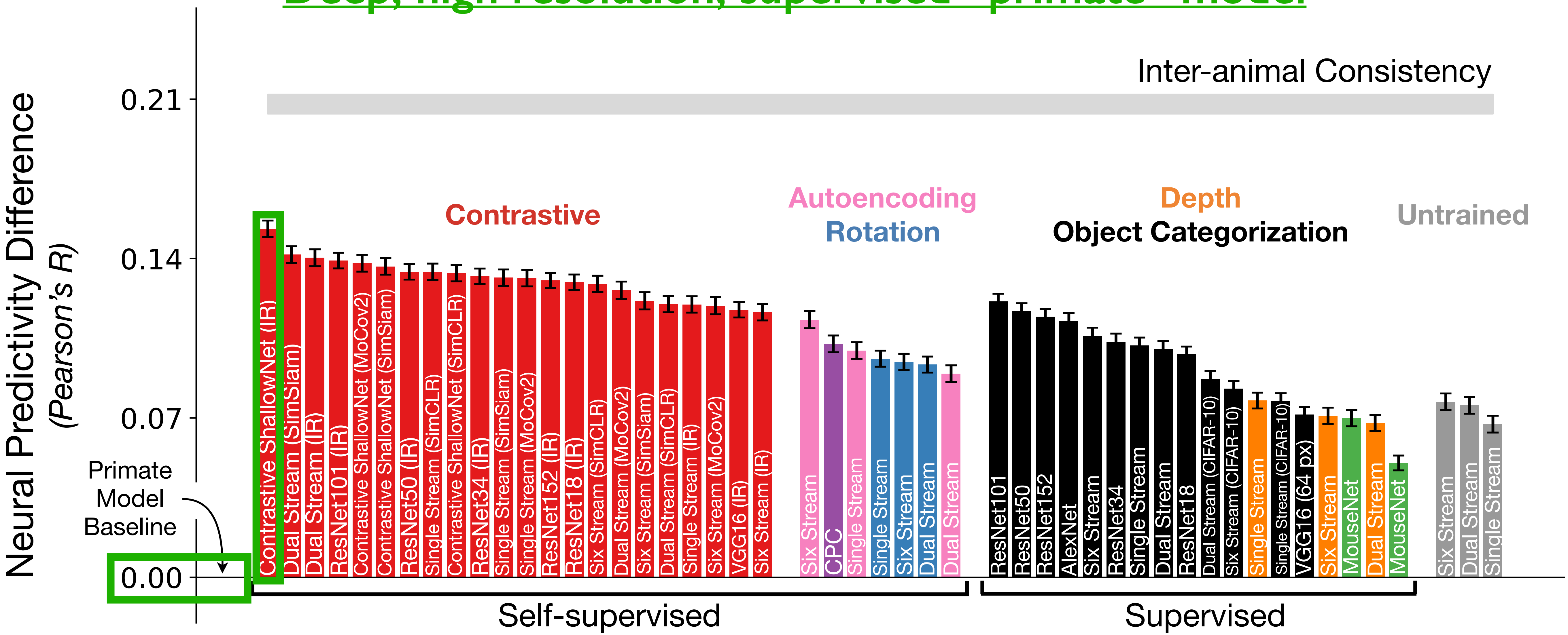




Substantially improving neural response predictivity of models of mouse visual cortex



Shallow, low resolution, unsupervised model
greatly improves over
Deep, high resolution, supervised "primate" model

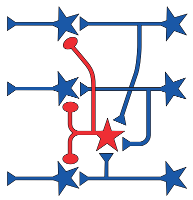


Improves neural predictivity from 56% to 90% of the data noise ceiling

Distilling Constraints: Circuit

A *= architecture class*

1. *"Circuit"*



shallow

~~deep~~

T *= task loss*

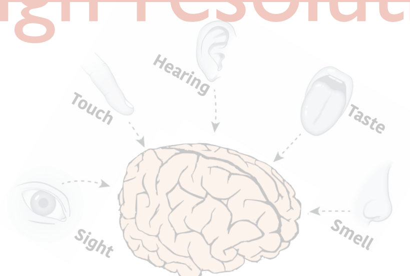
3. *"Ecological niche/behavior"*

unsupervised

~~supervised~~

low resolution

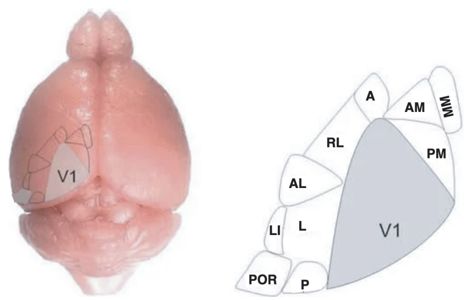
~~high resolution~~

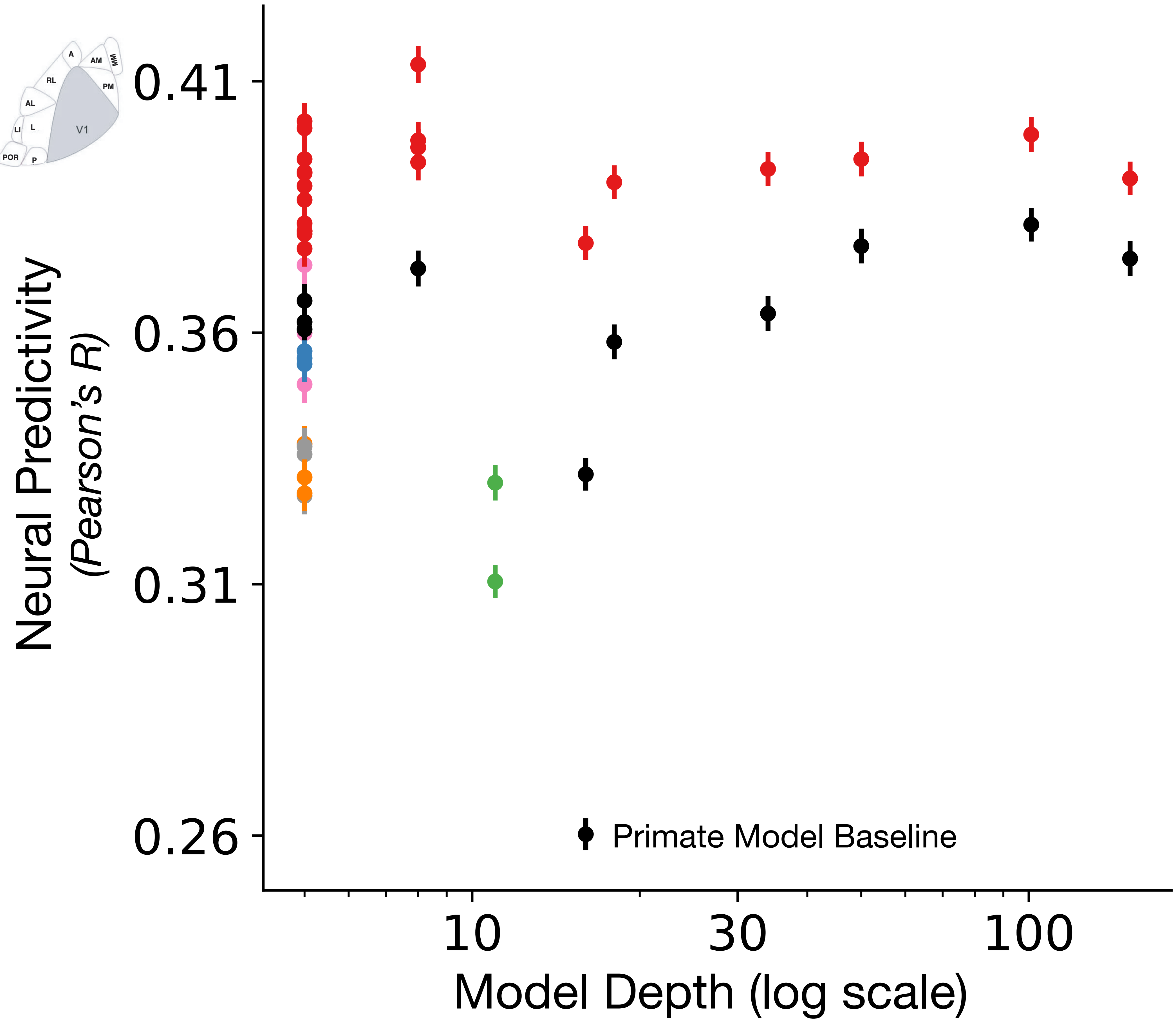


2. *"Environment"*

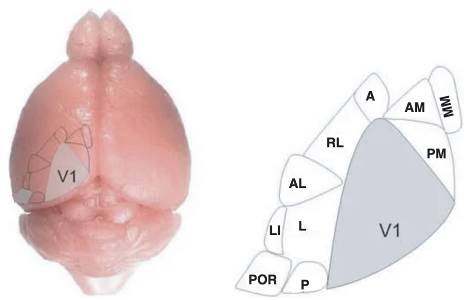
D *= data stream*

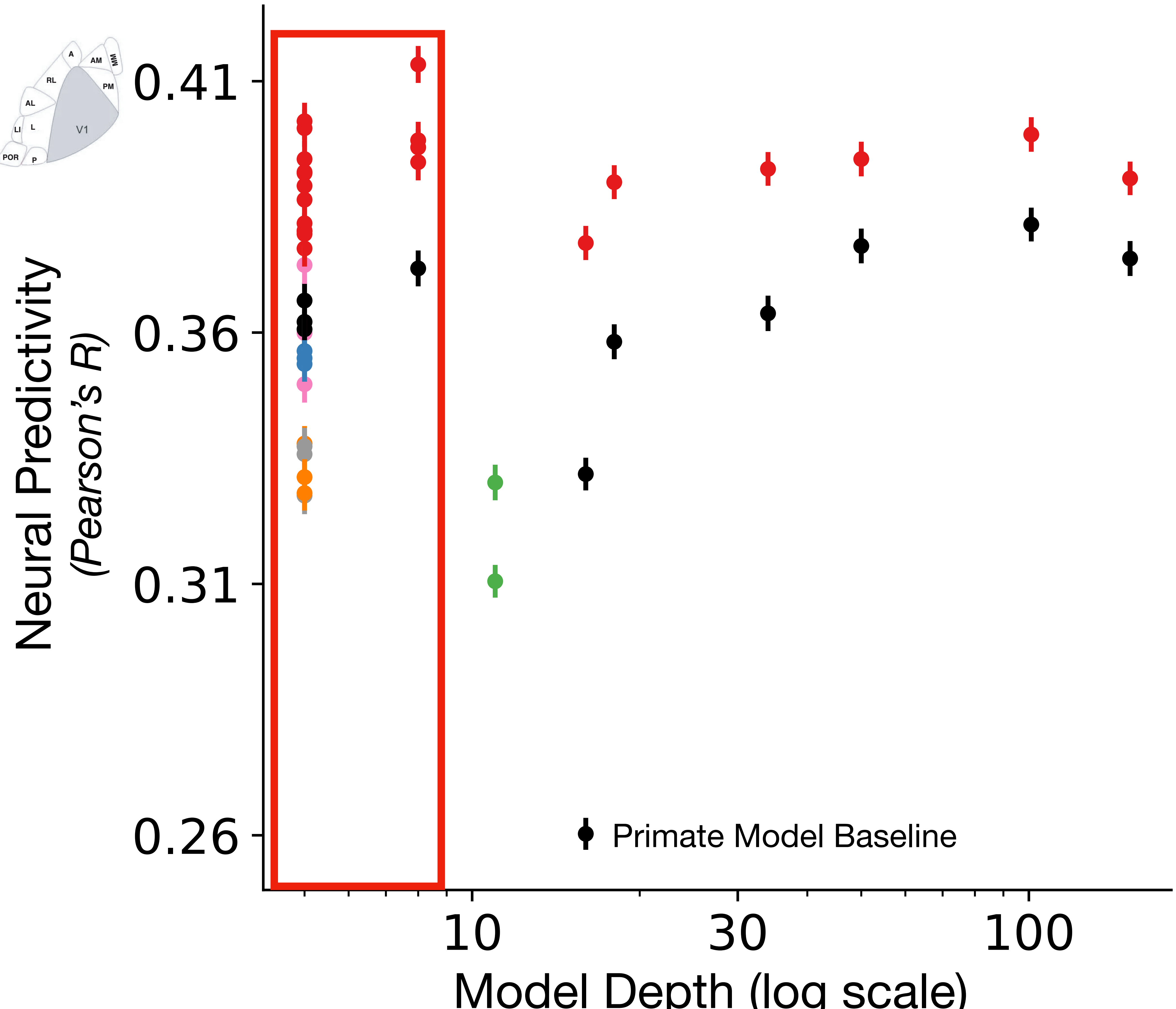
Architectural Choices



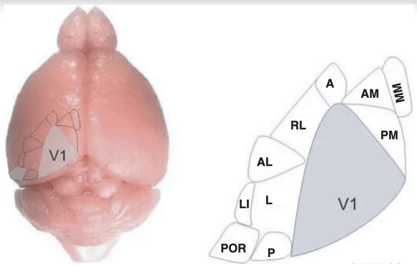


Architectural Choices: Shallower models suffice

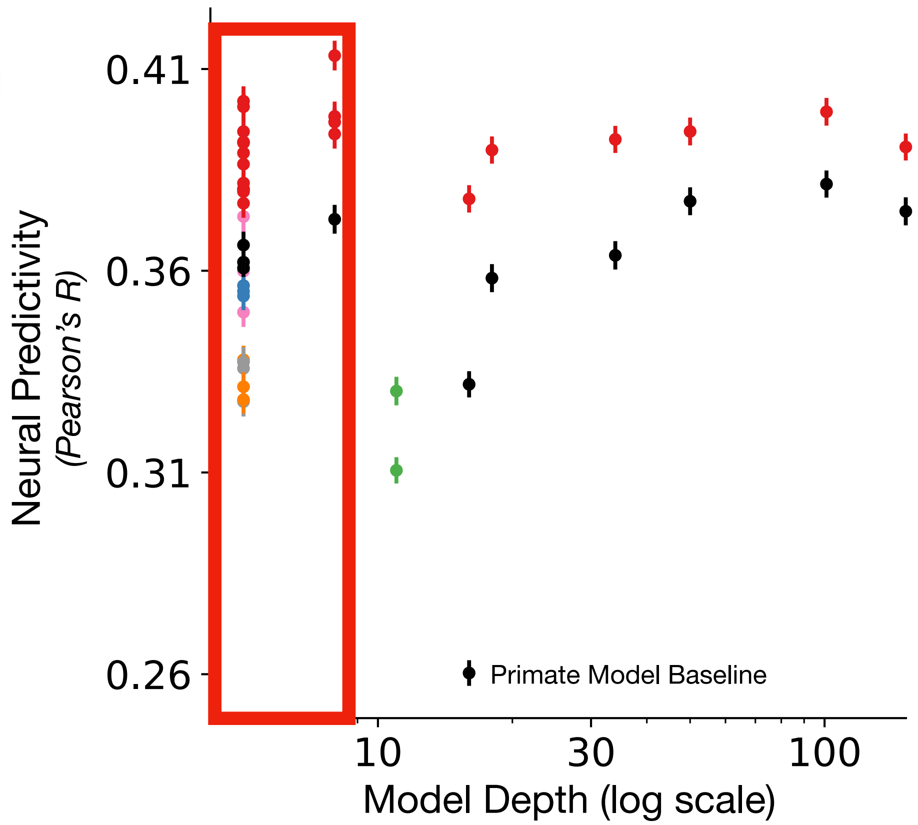


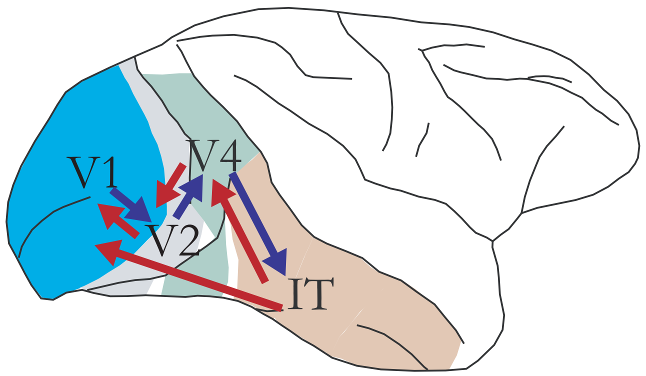


…unlike in primates!



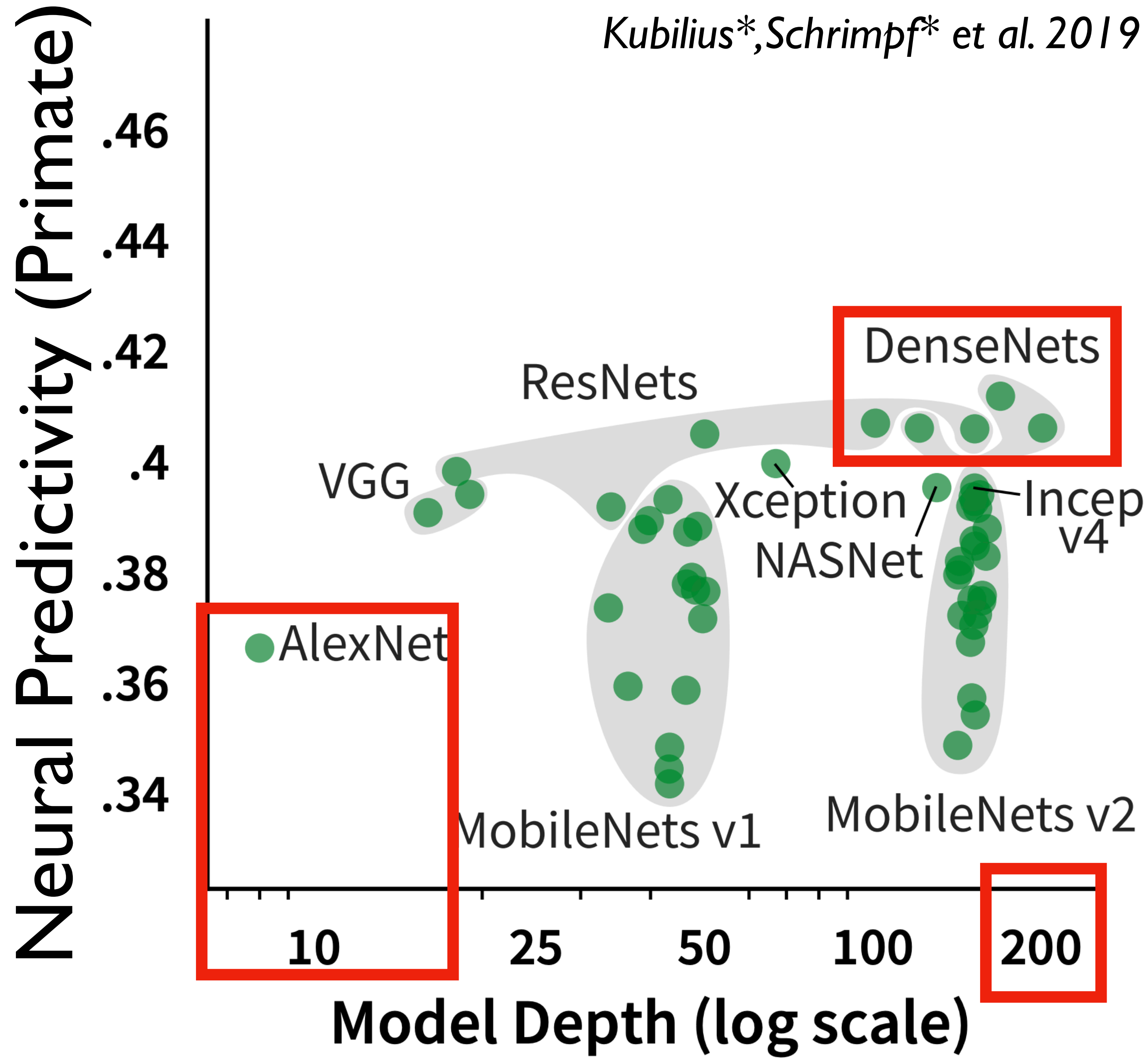
Mouse



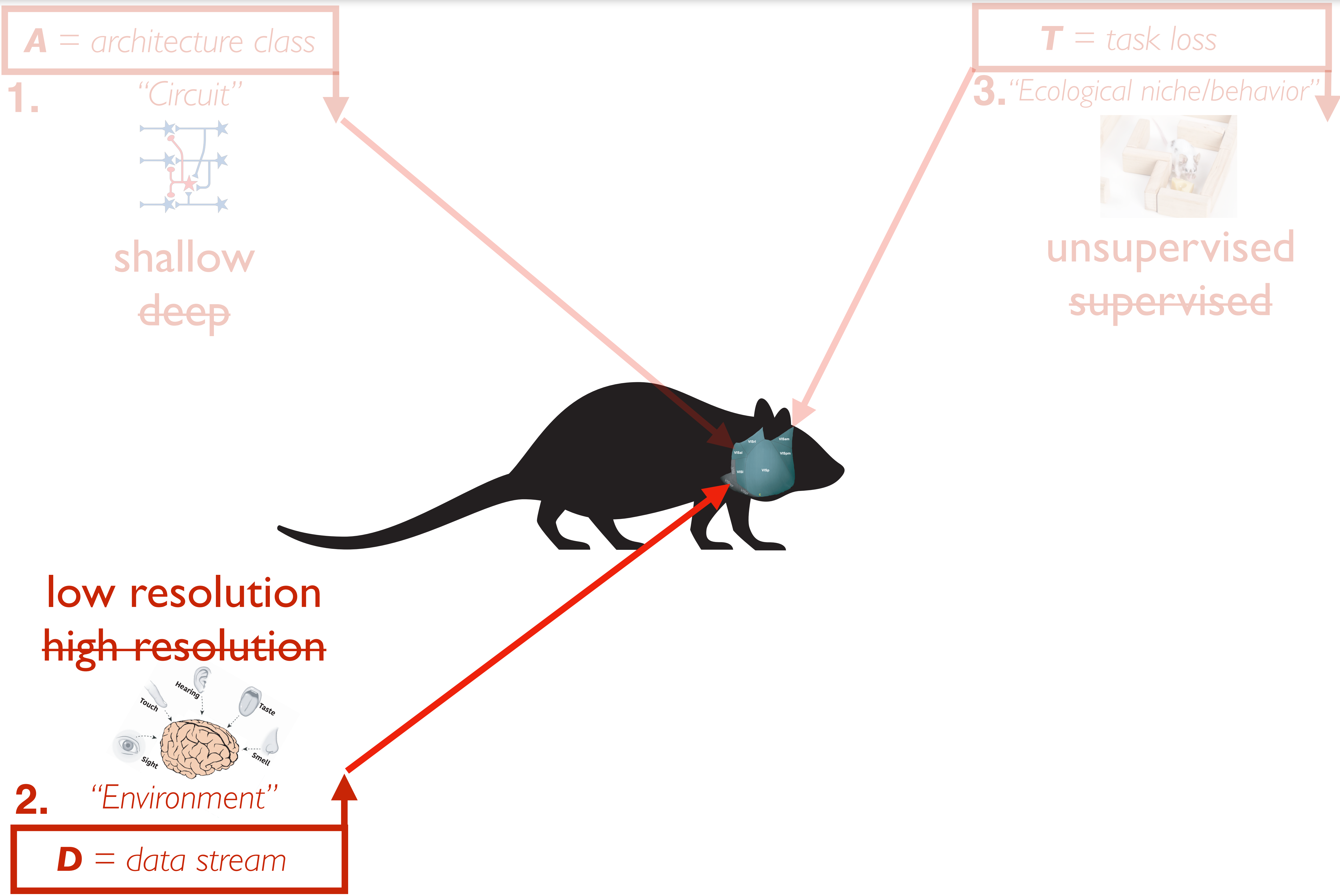


Unlike in primates!

Kubilius,Schrimpf* et al. 2019*

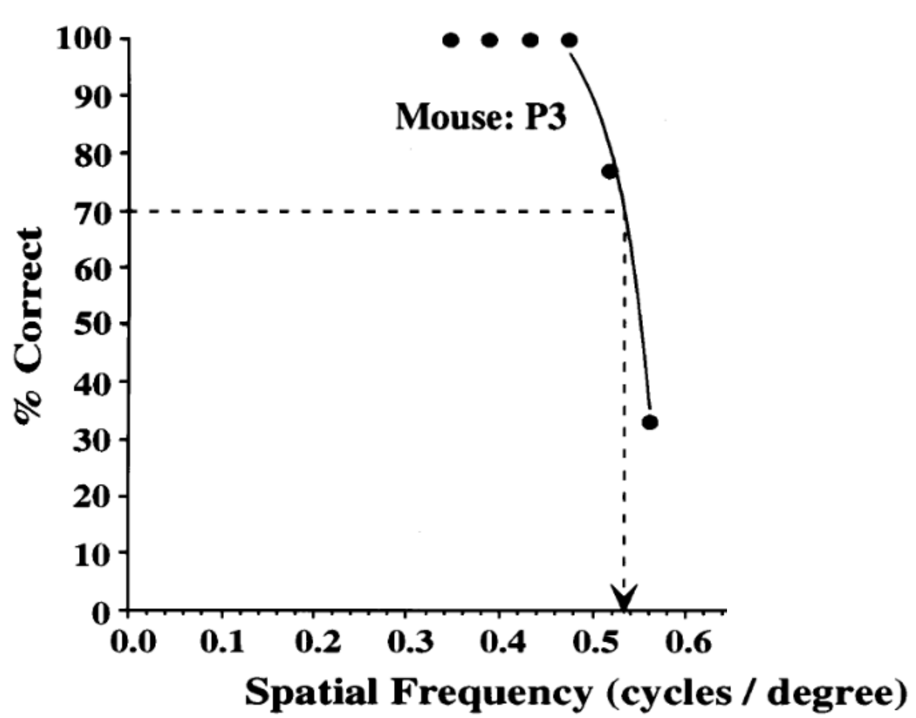


Distilling Constraints: Environment

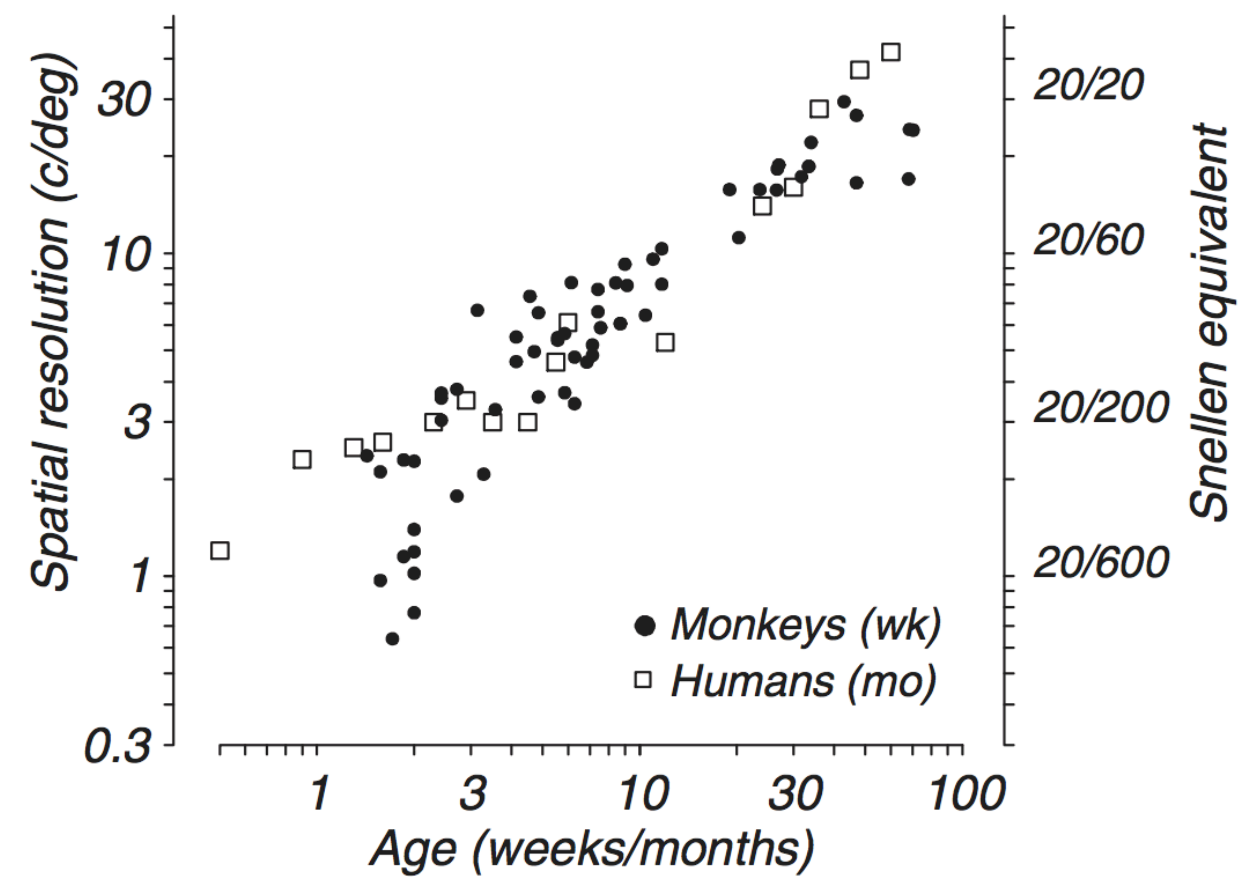


Distilling Constraints: Inputs

Mice



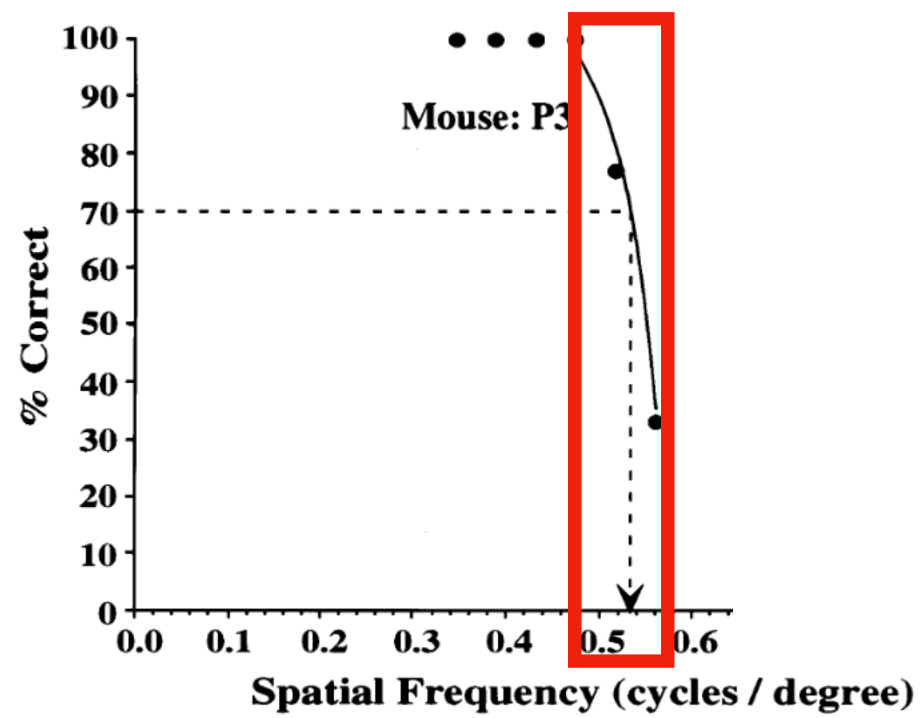
Primates



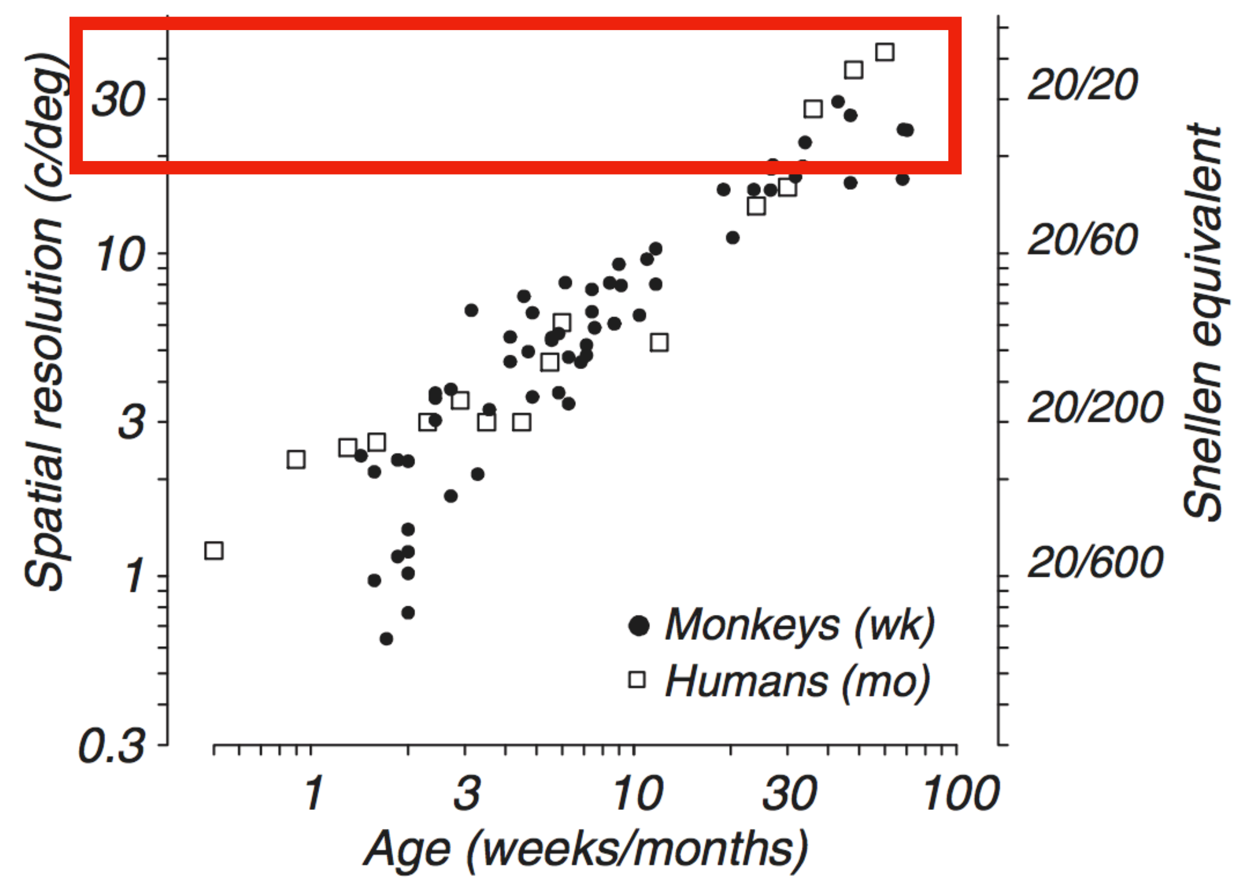
Prusky et al., 2000; Kiorpes, 2019

Distilling Constraints: Inputs

Mice

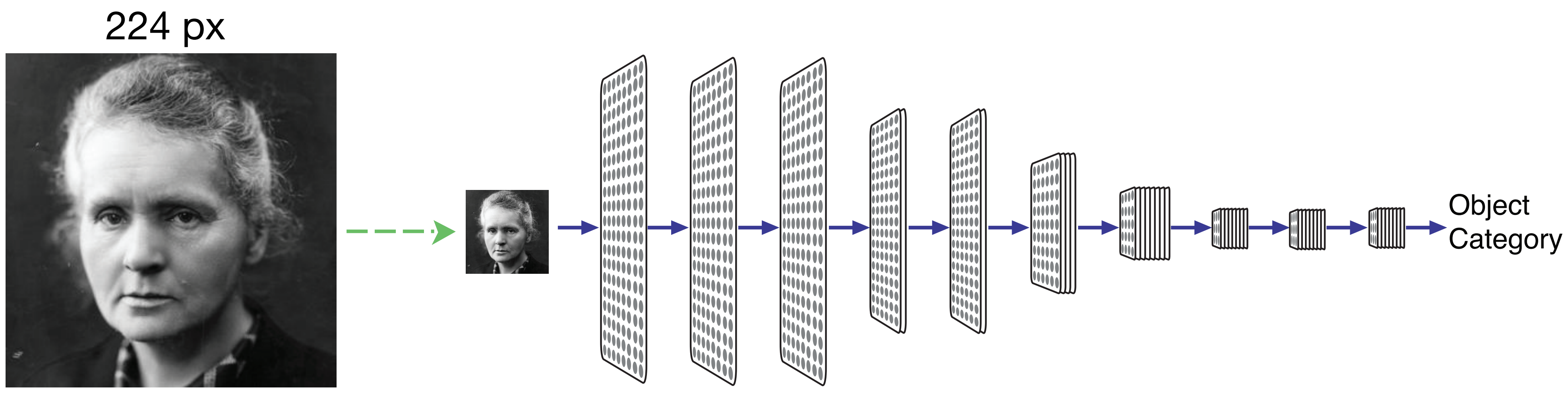


Primates

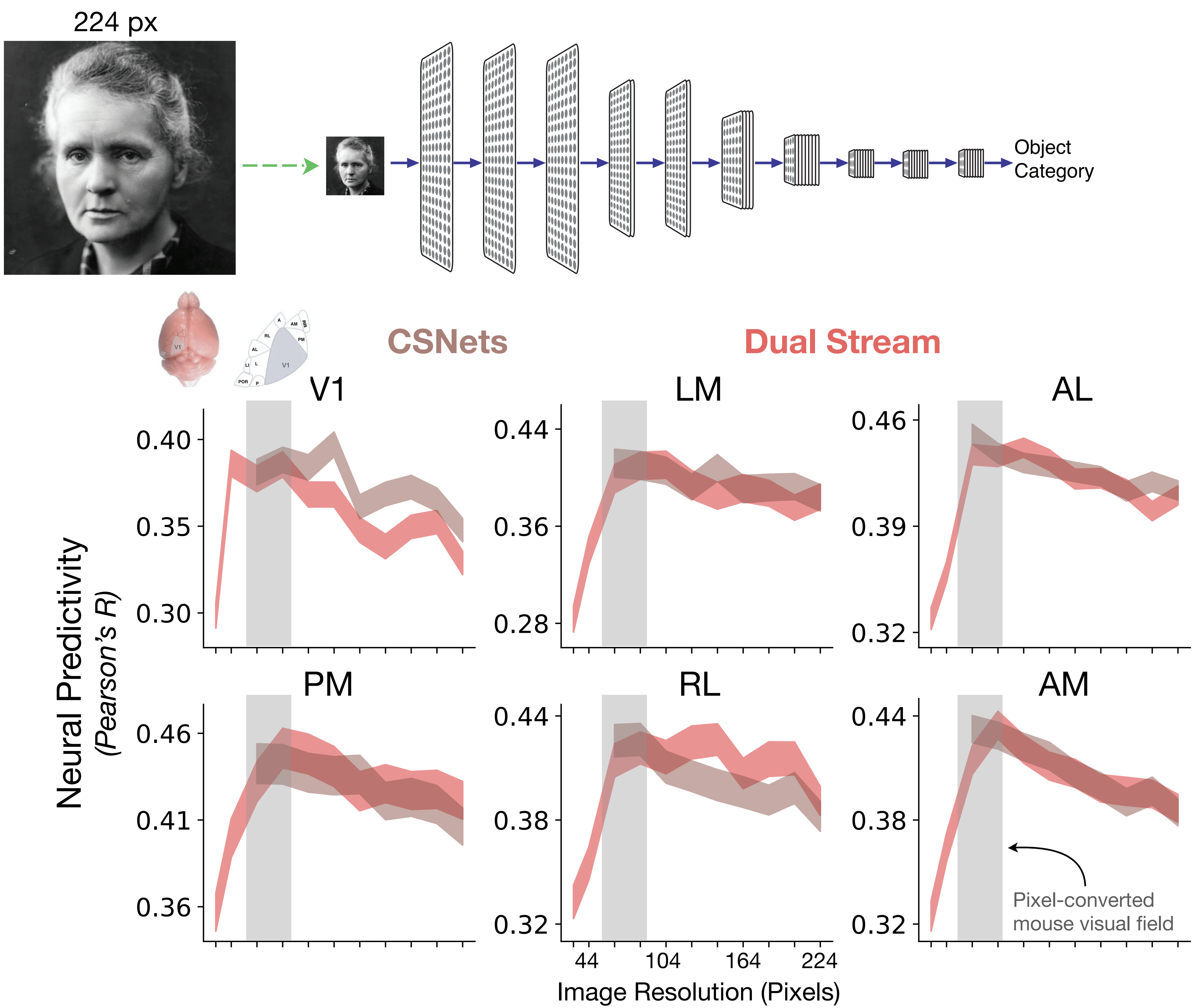


Prusky et al., 2000; Kiorpes, 2019

Distilling Constraints: Inputs

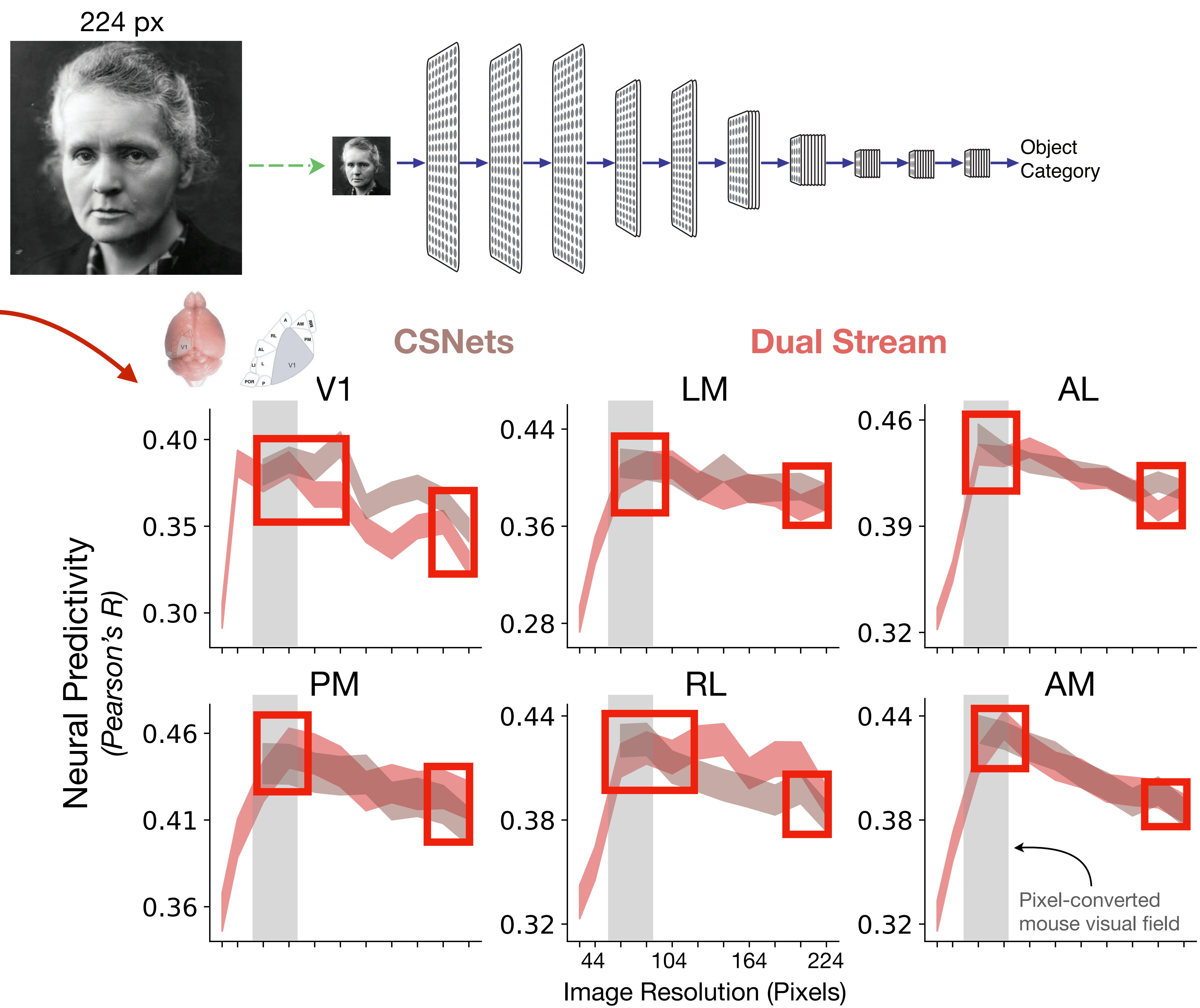


Distilling Constraints: Inputs



Lower resolution improves neural predictivity

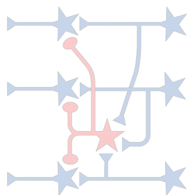
Reducing image resolution during training improves neural predictivity across visual areas



Distilling Constraints: Behavioral Goals

A = *architecture class*

1. *“Circuit”*



shallow

~~deep~~

T = *task loss*

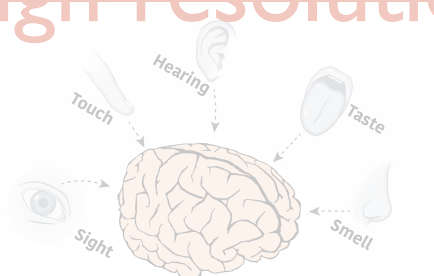
3. *“Ecological niche/behavior”*

unsupervised

~~supervised~~

low resolution

~~high resolution~~



2. *“Environment”*

D = *data stream*

Distilling Constraints: Behavioral Goals

Supervised Losses

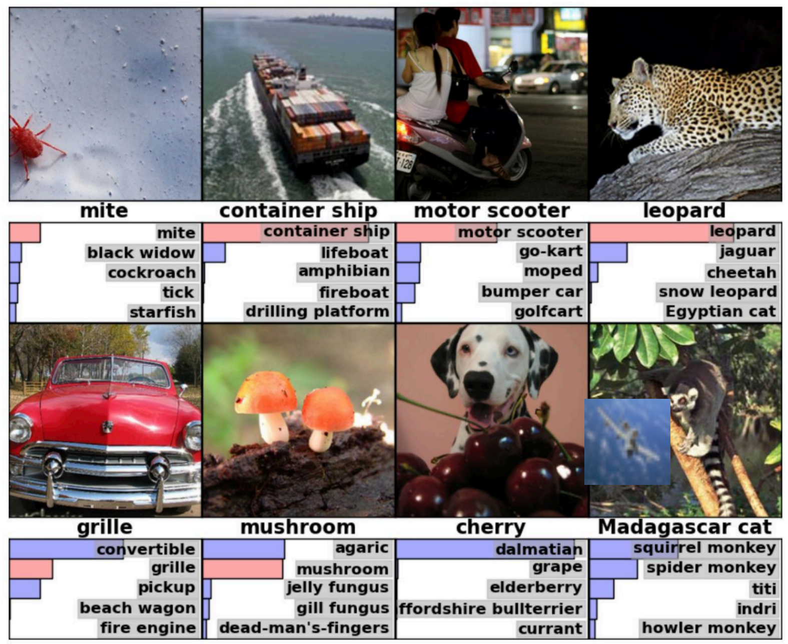
Distilling Constraints: Behavioral Goals

Supervised Losses

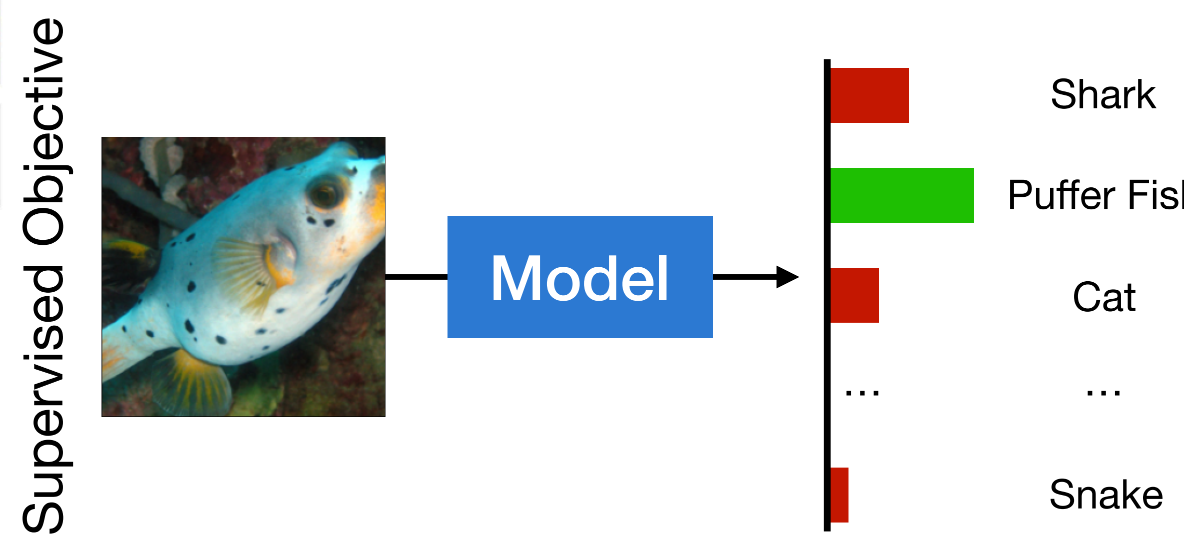
ImageNet Challenge

IMAGENET

- 1,000 object classes (categories).
- Images:
  - 1.2 M train
  - 100k test.



Most Supervision (1000 classes)



Typical setting: supervision with (1000) category labels ...but is very "unnatural" for mice!

Both the type and number of categories is unrealistic for mice

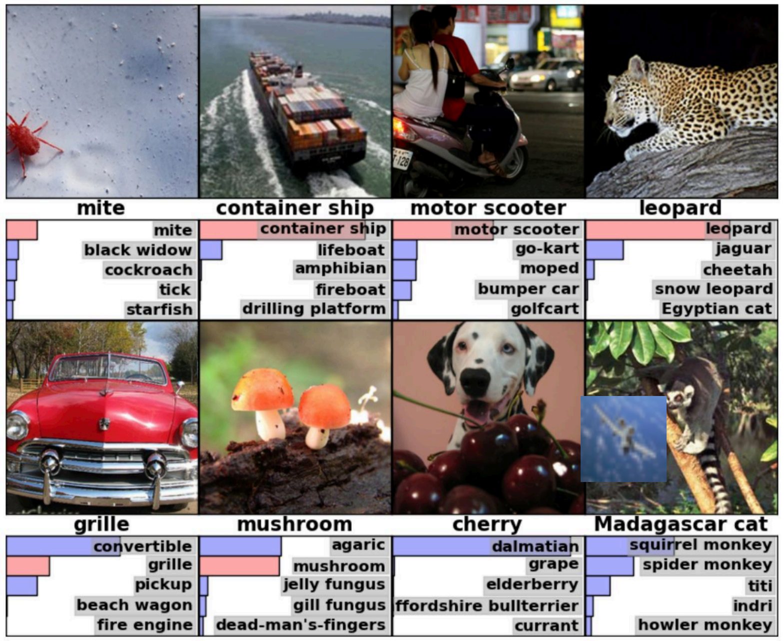
Distilling Constraints: Behavioral Goals

Supervised Losses

ImageNet Challenge

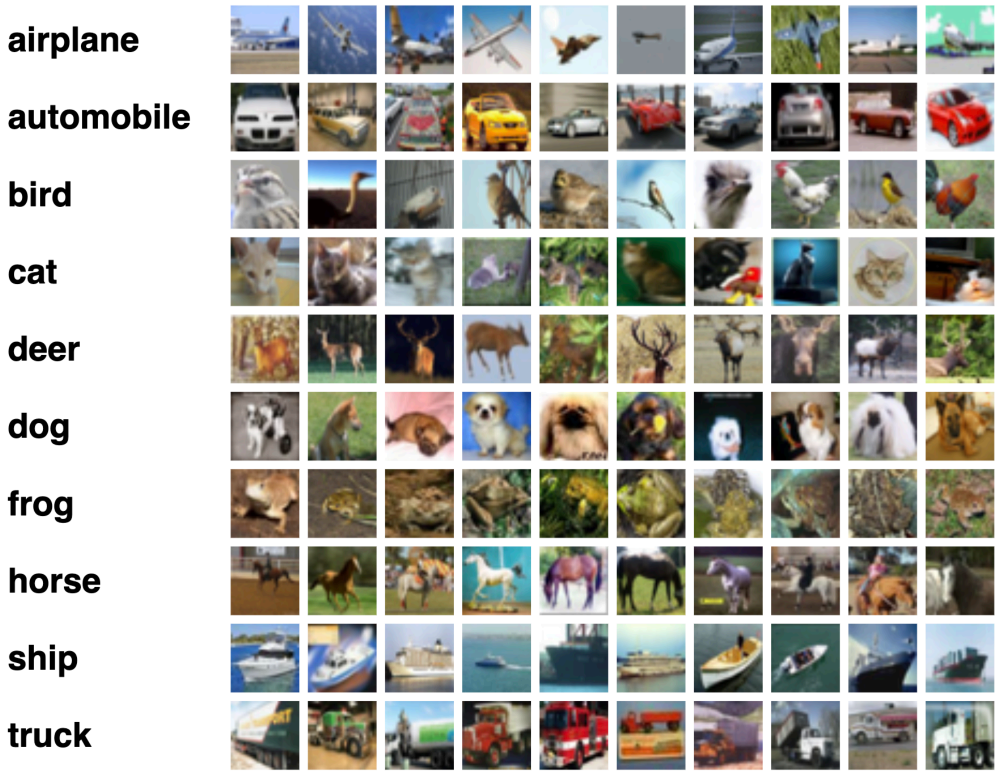
IMAGENET

- 1,000 object classes (categories).
- Images:
  - 1.2 M train
  - 100k test.



Most Supervision
(1000 classes)

CIFAR-10



Less Supervision
(10 object classes)

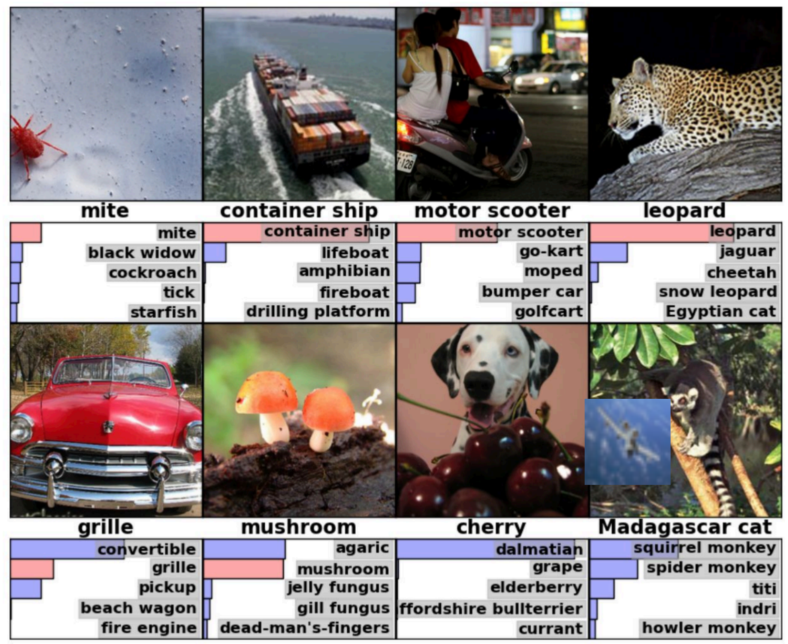
Distilling Constraints: Behavioral Goals

Supervised Losses

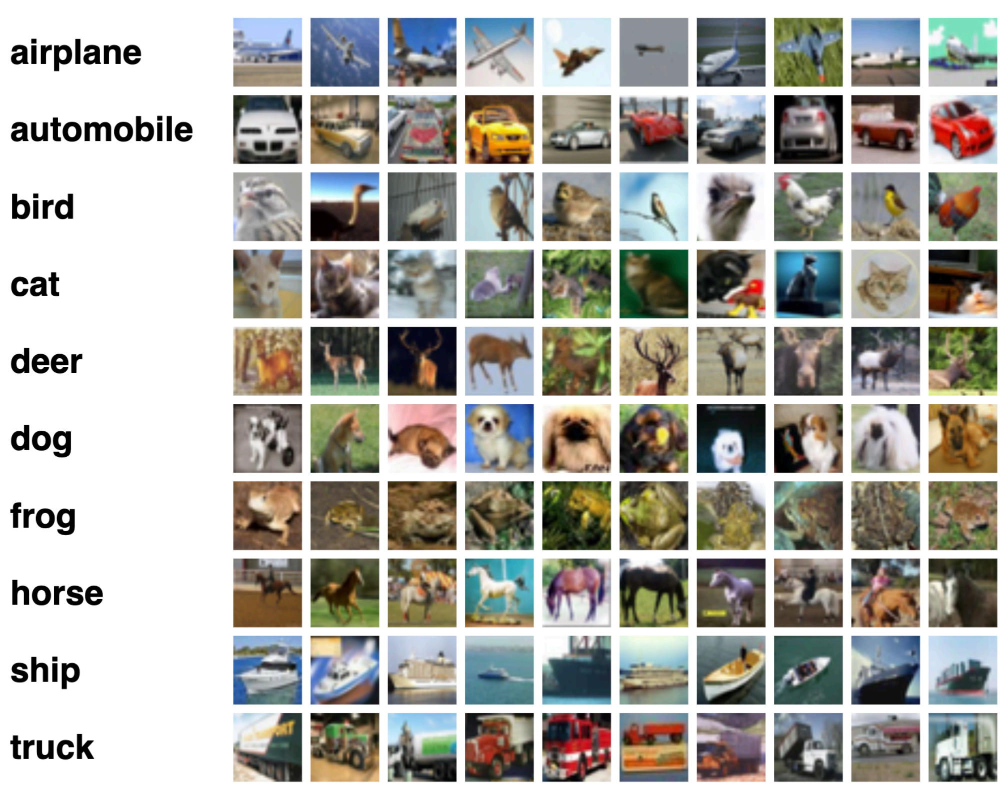
ImageNet Challenge

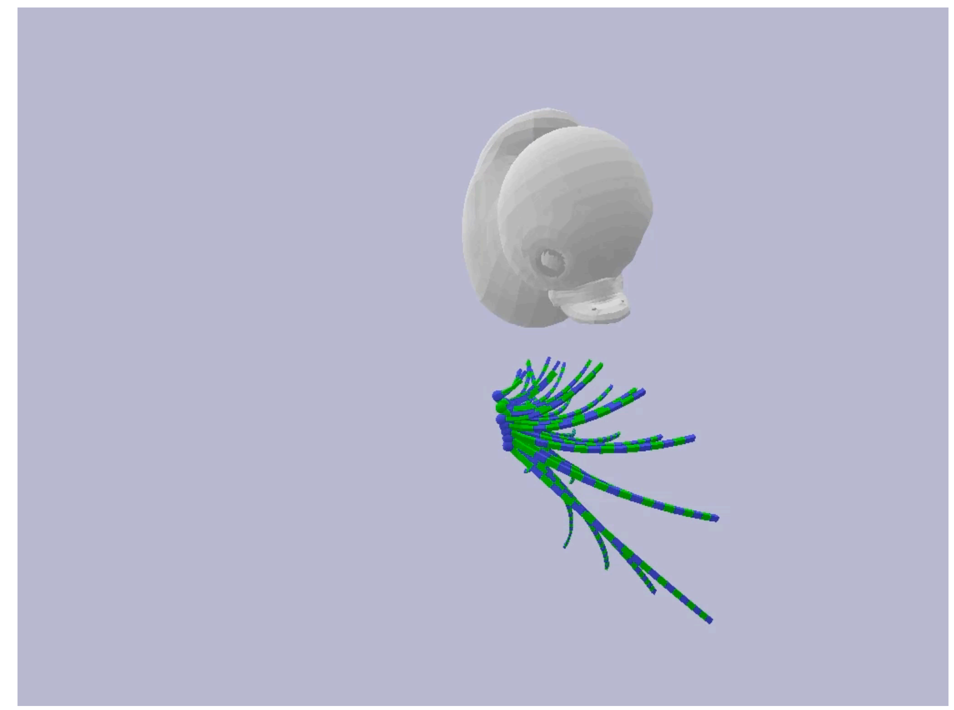
IMAGENET

- 1,000 object classes (categories).
- Images:
  - 1.2 M train
  - 100k test.



CIFAR-10





Depth Map Prediction
(Visual proxy for whisking)

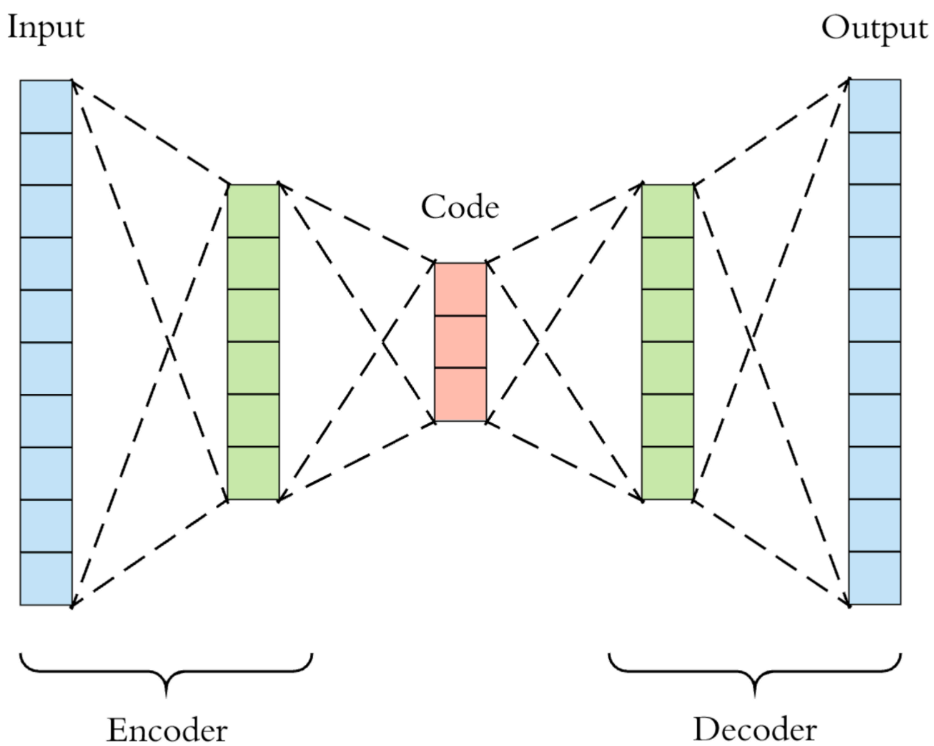
PBRNet (500K images)

Distilling Constraints: Behavioral Goals

Unsupervised Losses

Unsupervised Losses

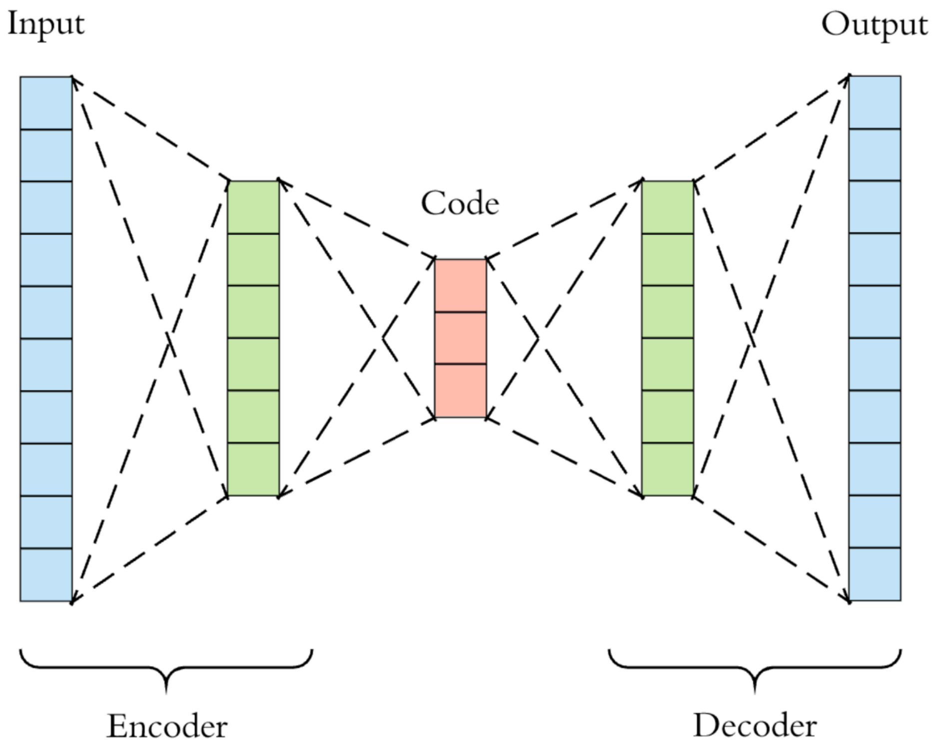
Sparse Autoencoding

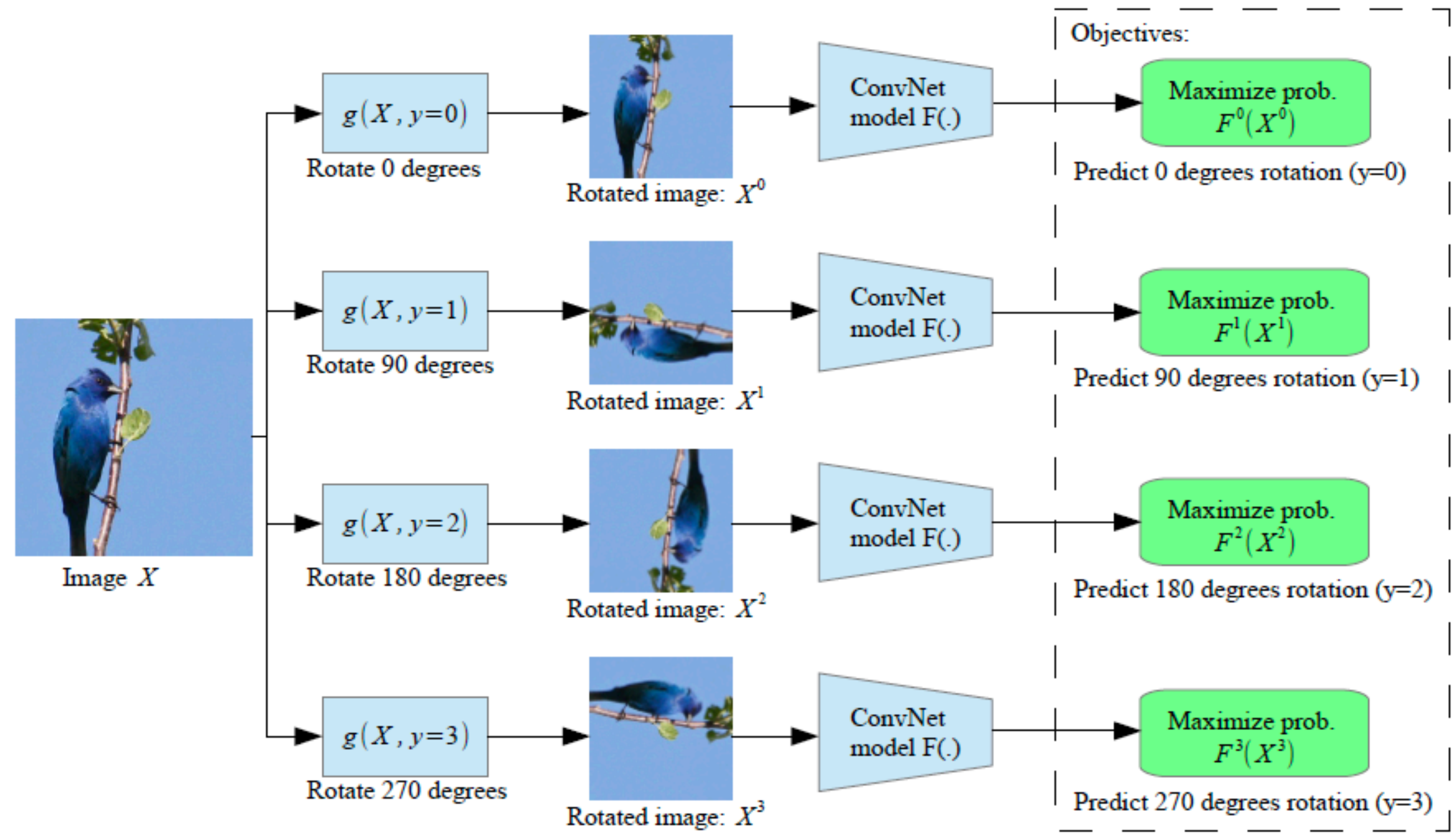


Distilling Constraints: Behavioral Goals

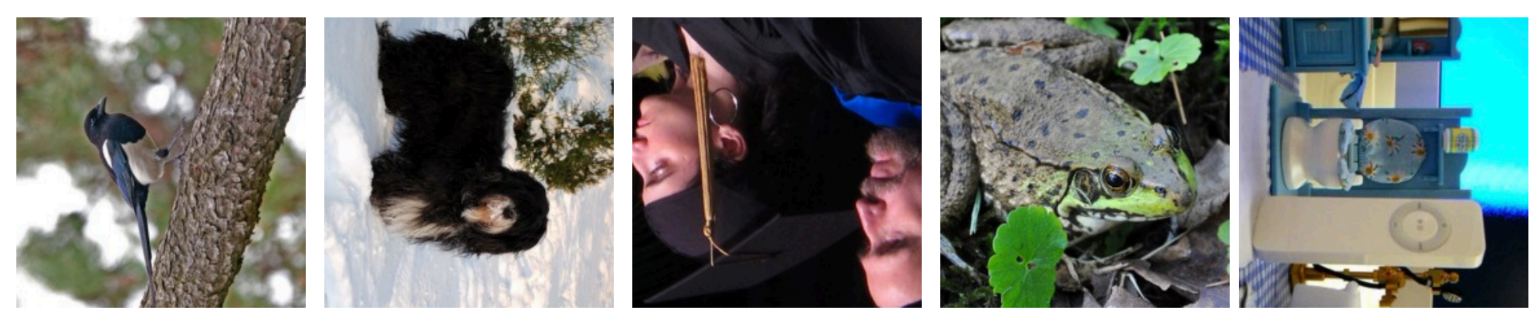
Unsupervised Losses

Sparse Autoencoding





Predict Image Rotations (RotNet, Gidaris et al. 2018)



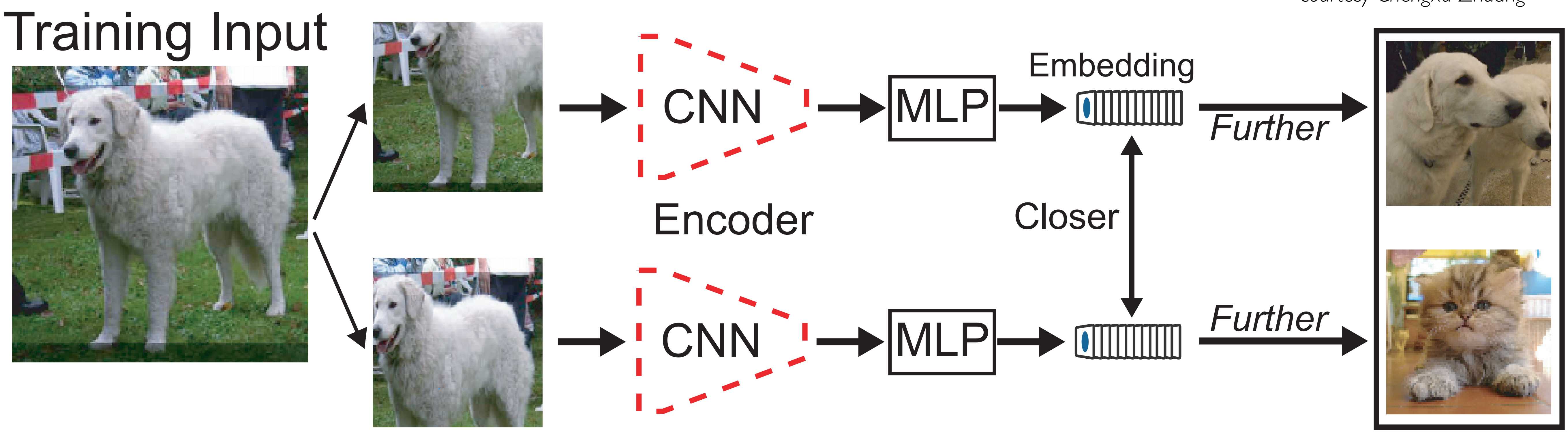
90° rotation | 270° rotation | 180° rotation | 0° rotation | 270° rotation

Distilling Constraints: Behavioral Goals

Unsupervised Losses

Contrastive Objectives

courtesy Chengxu Zhuang



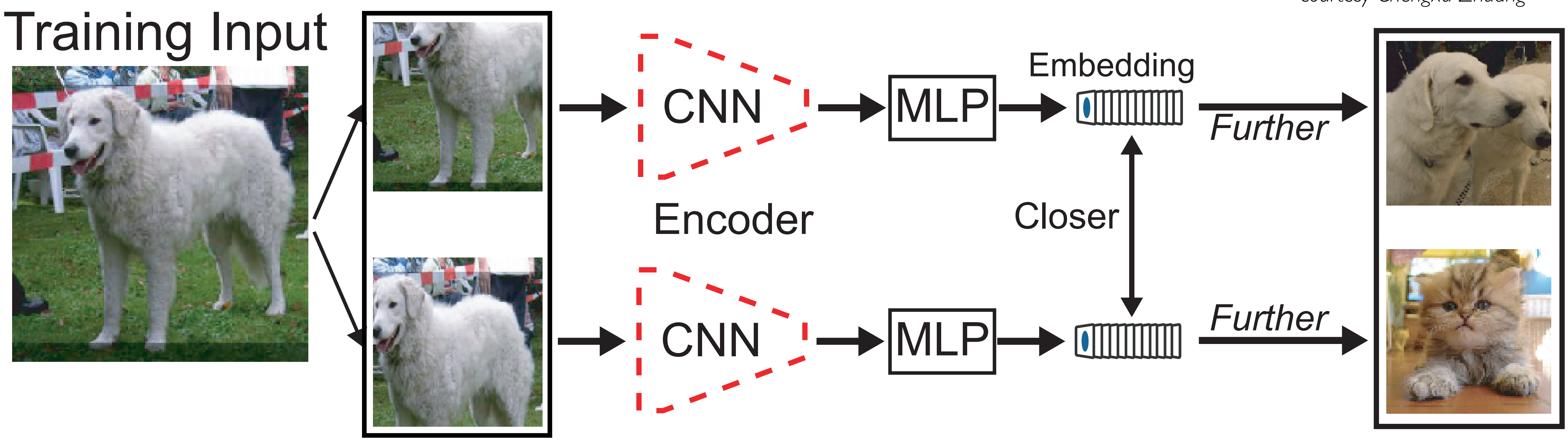
CNN: Convolutional Neural Network, MLP: Multi-Layer Perceptron

Distilling Constraints: Behavioral Goals

Unsupervised Losses

<u>Contrastive Objectives</u>

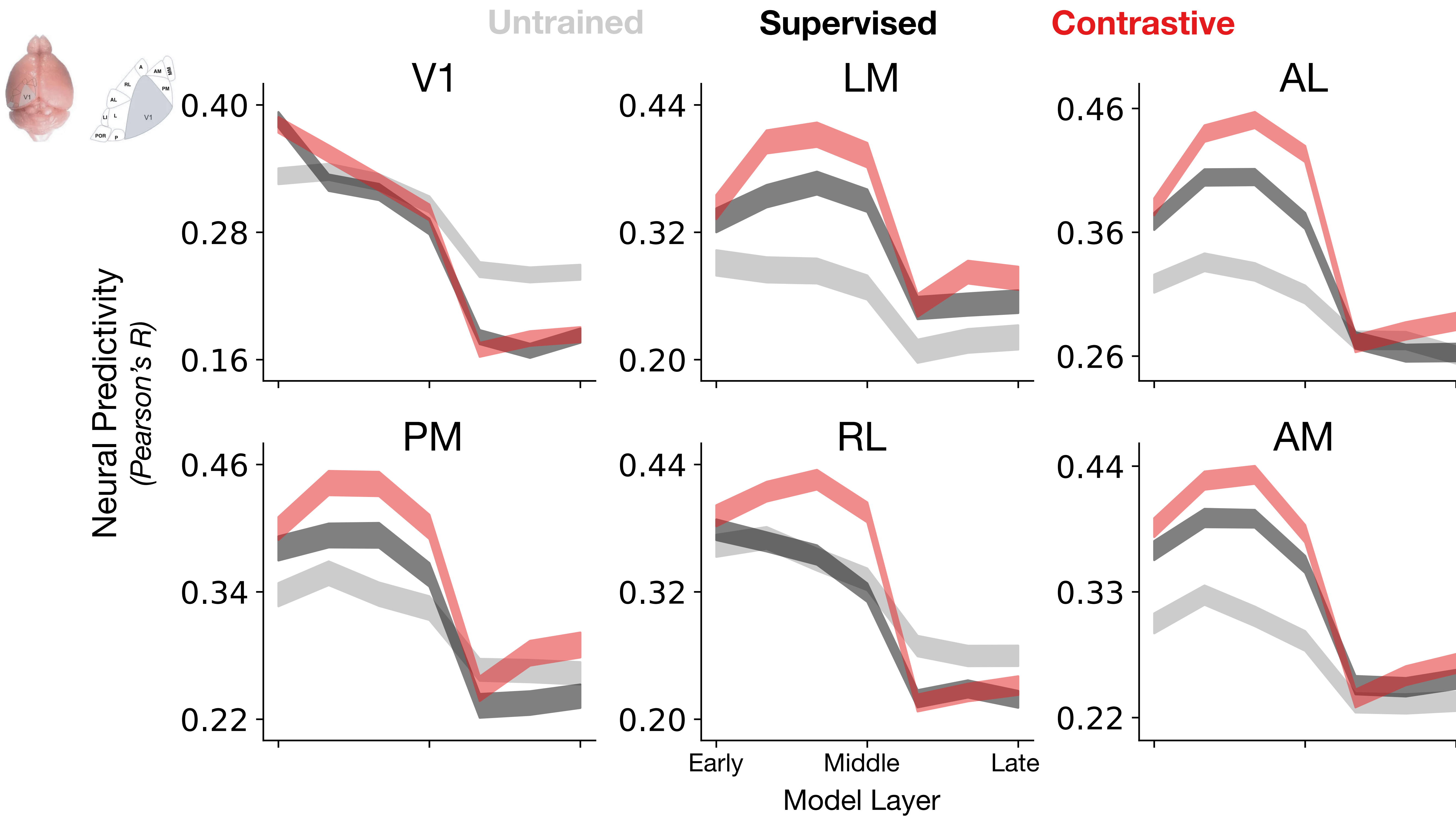
courtesy Chengxu Zhuang



CNN: Convolutional Neural Network, MLP: Multi-Layer Perceptron

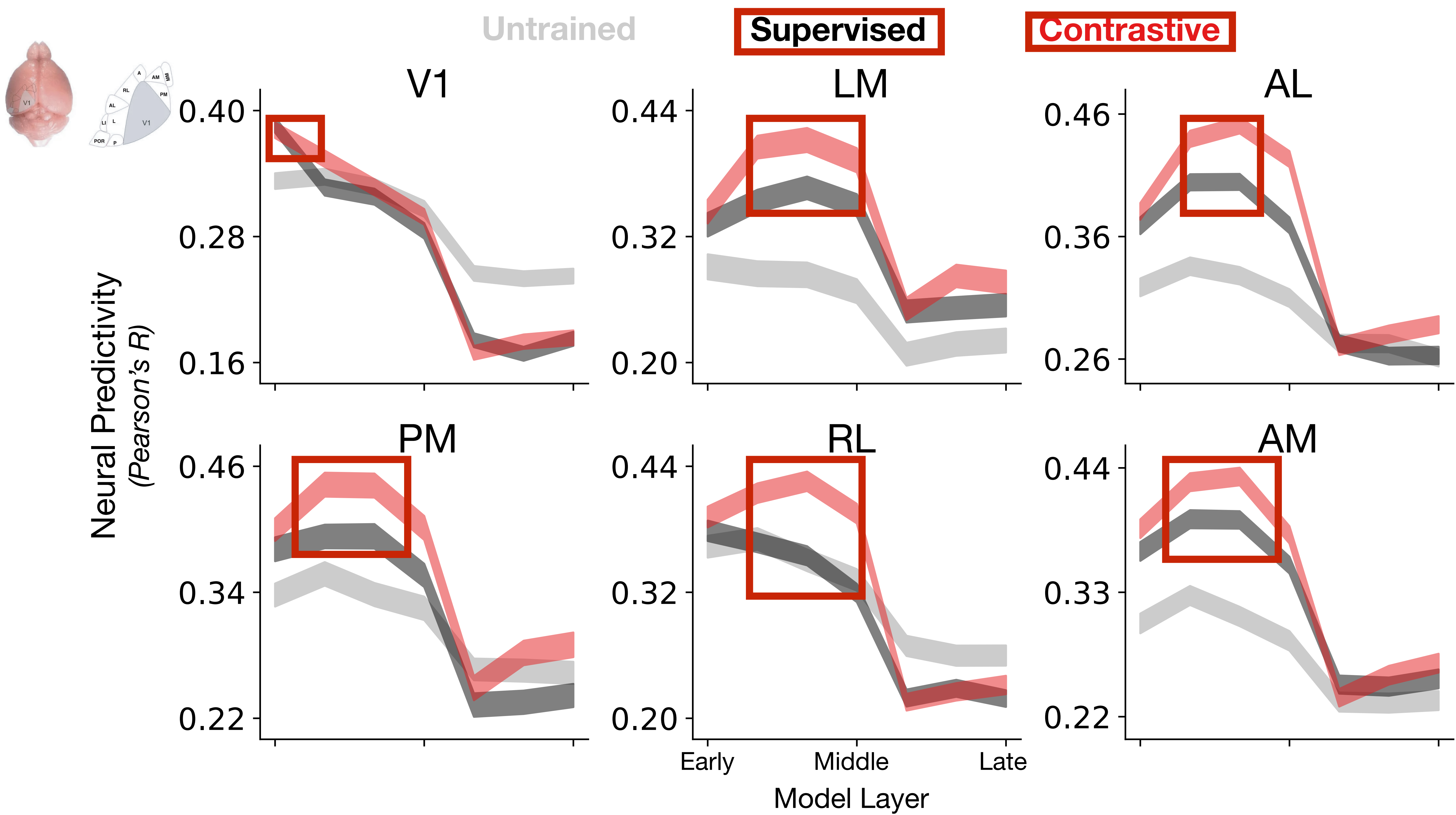
Make the embeddings of <u>different</u> views of the *same* image to be similar, while pushing them apart from different images

Unsupervised > Supervised Losses



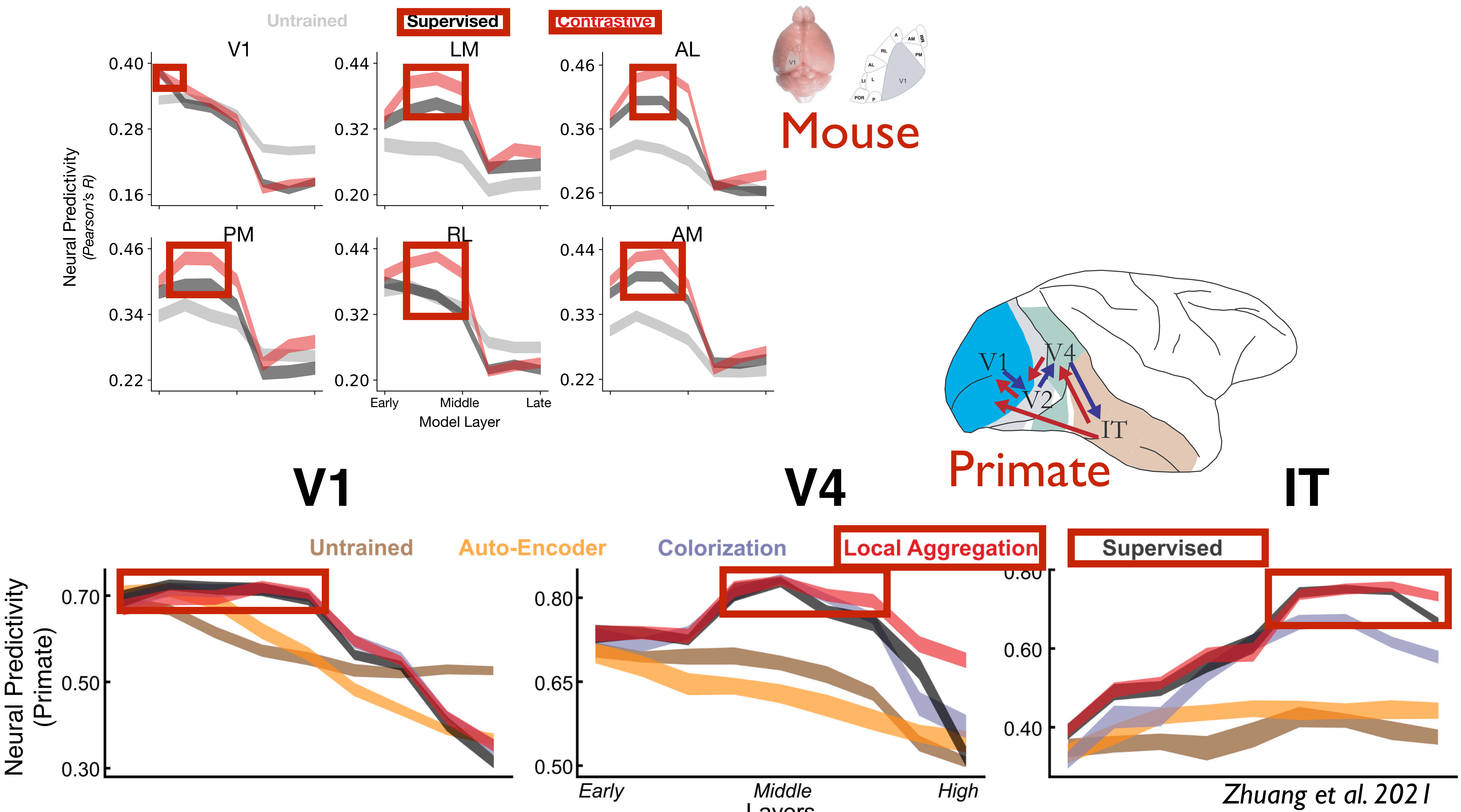
Unsupervised > Supervised Losses

Using an unsupervised, contrastive objective function improves neural predictivity (*best* architecture & data stream fixed)



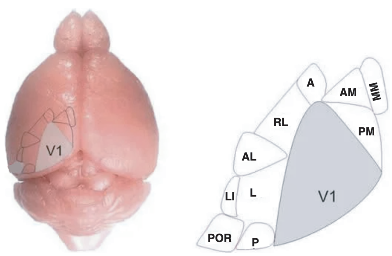
Unsupervised > Supervised Losses

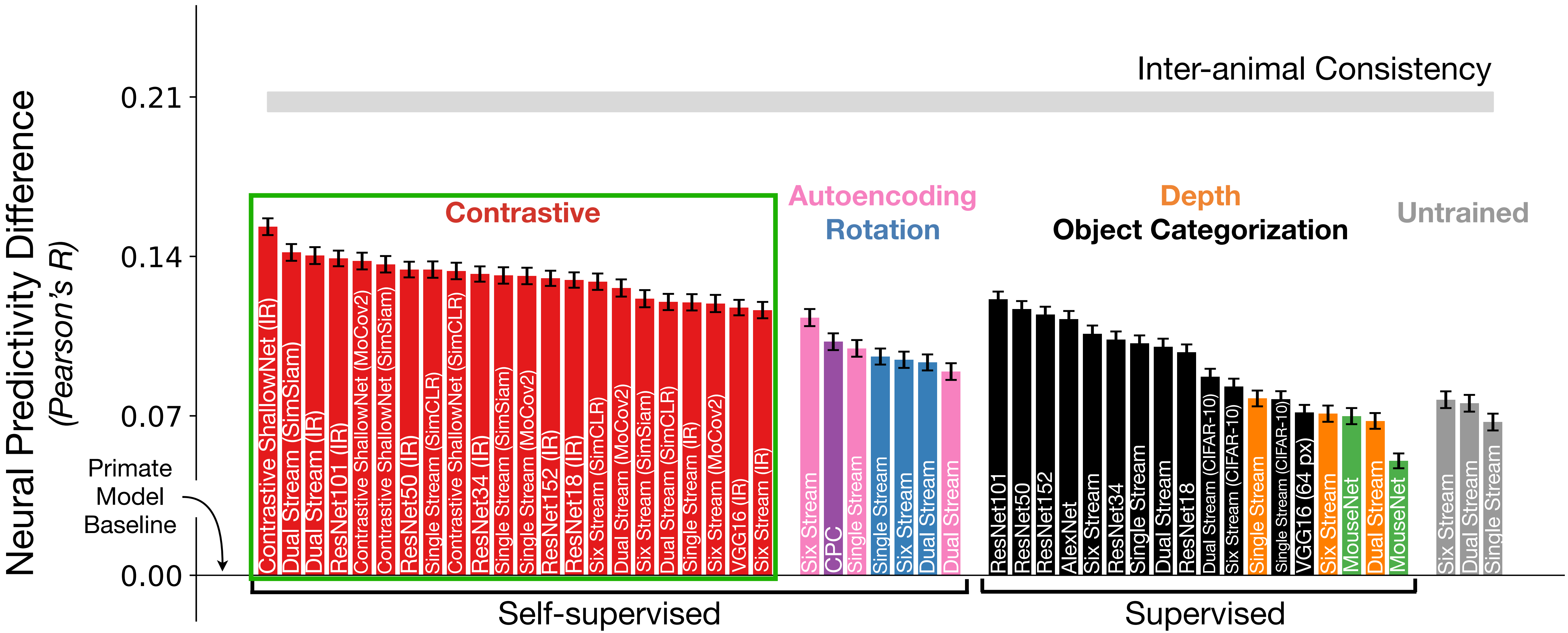
Unlike in primates where contrastive matches supervised!



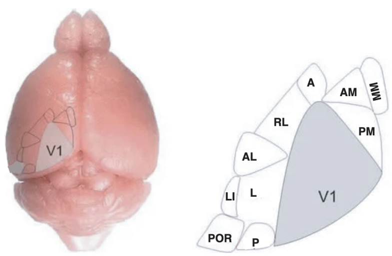
Zhuang et al. 2021

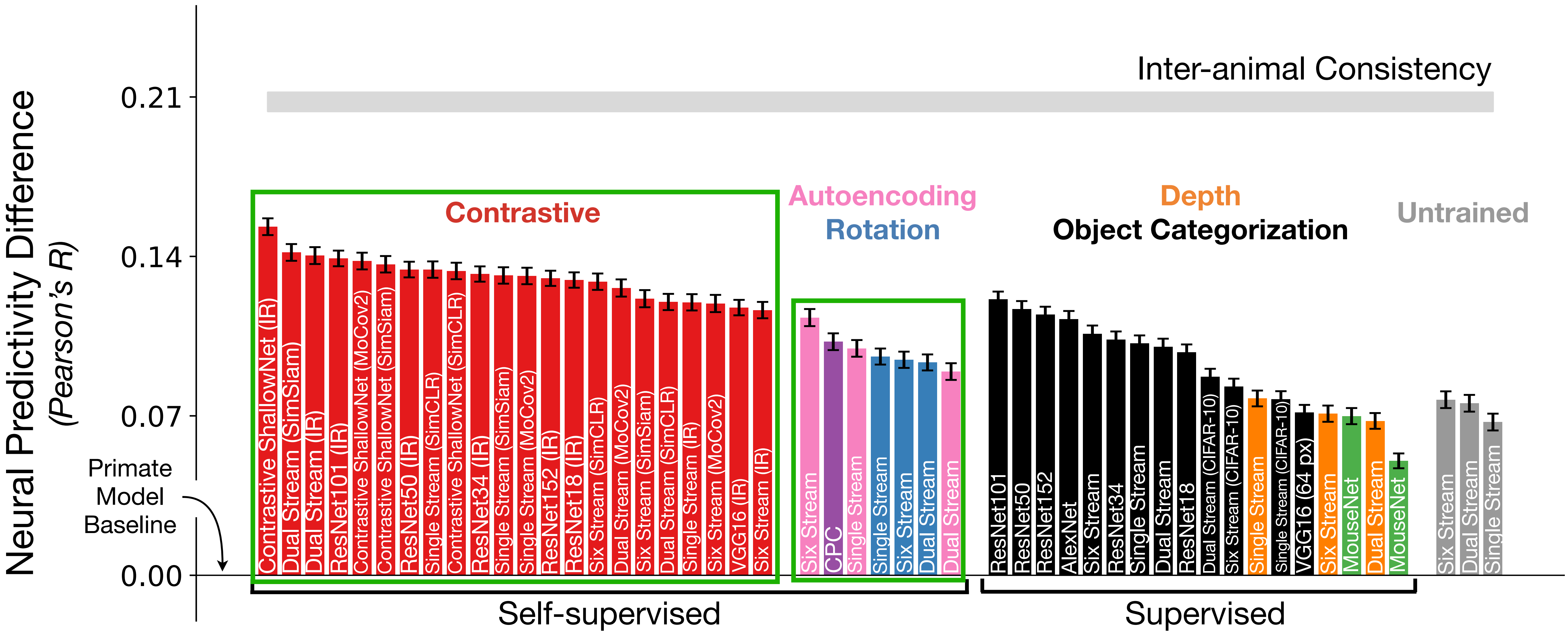
Contrastive losses are overall the best *unsupervised* loss





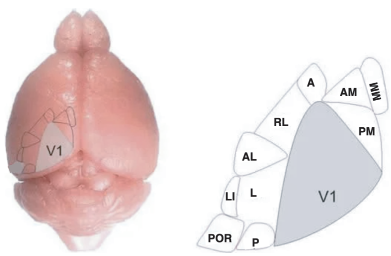
Contrastive losses are overall the best *unsupervised* loss

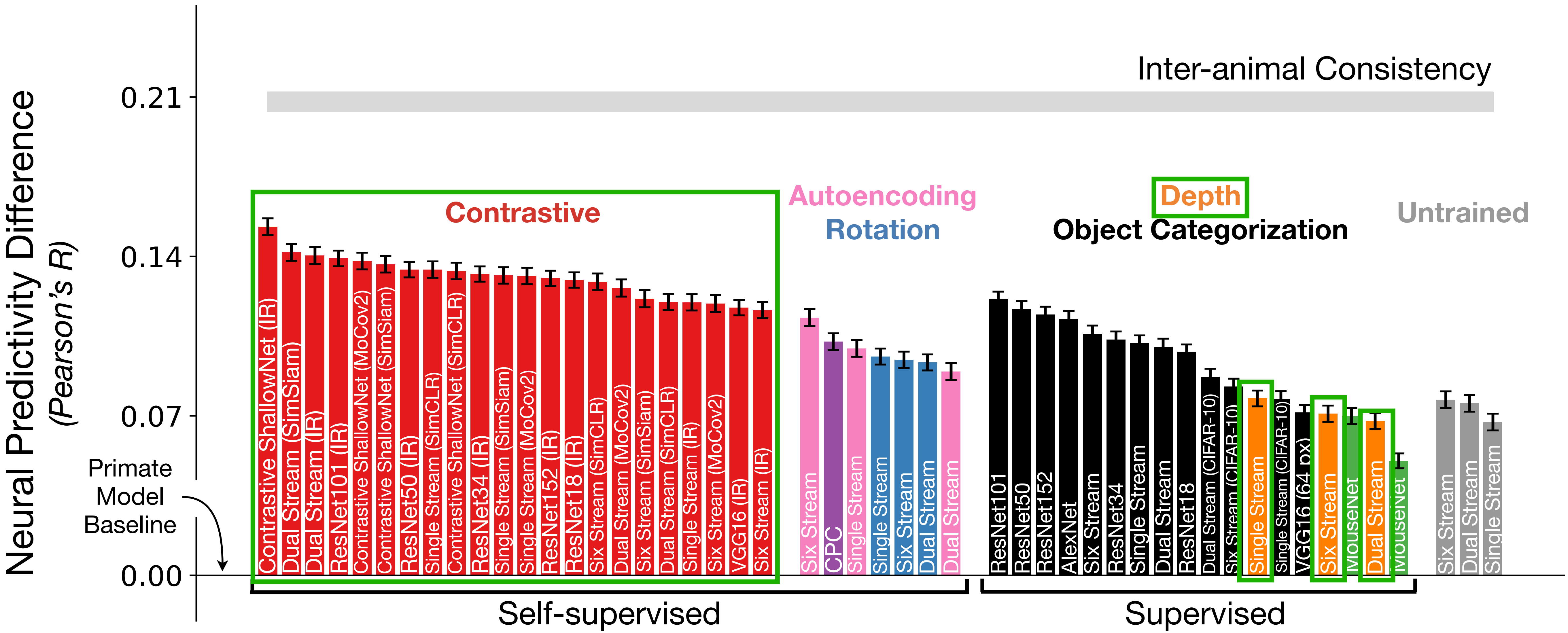




Contrastive objectives are overall the best compared to other *unsupervised* alternatives

Unsupervised losses outperform behaviorally-driven loss

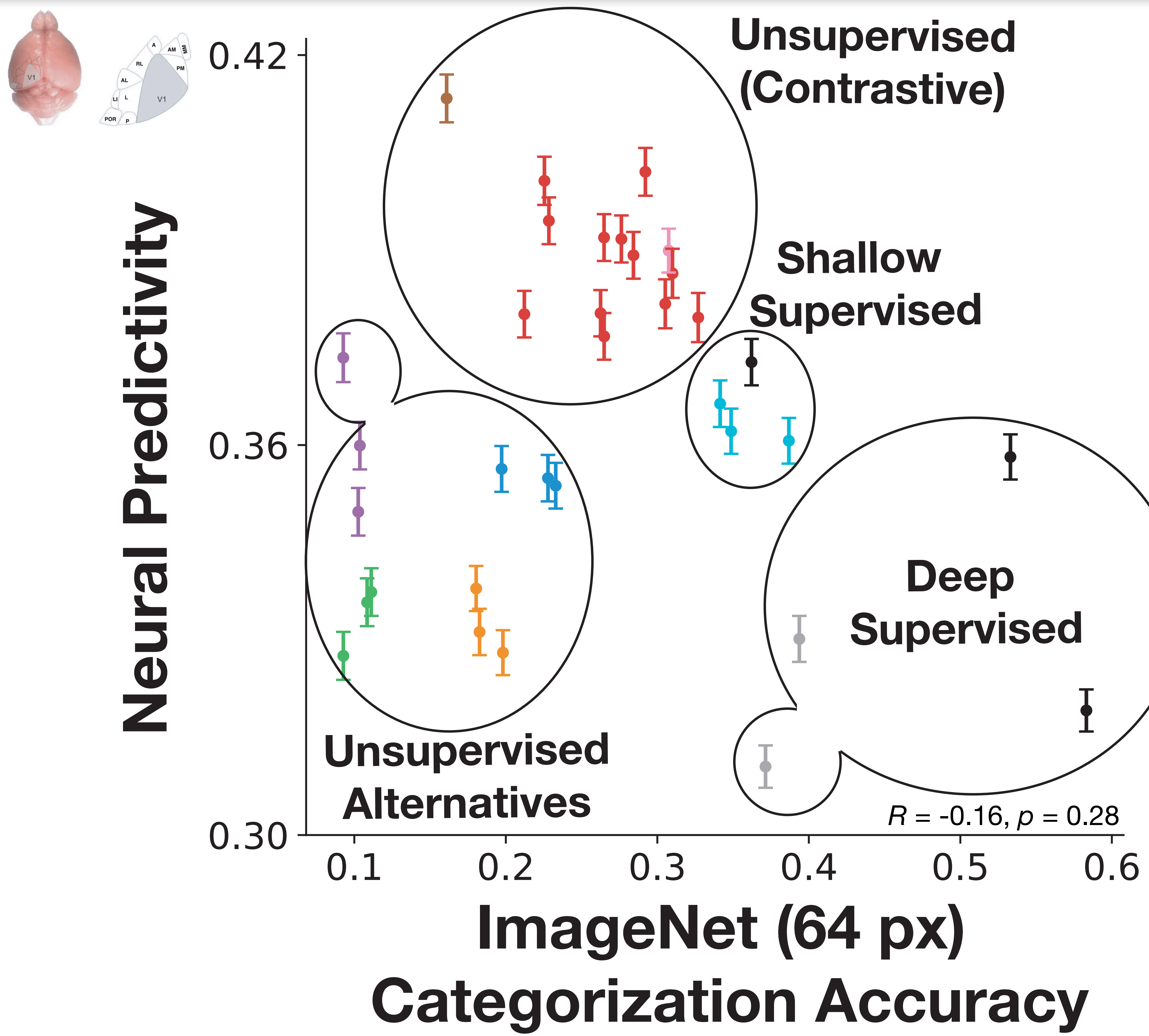




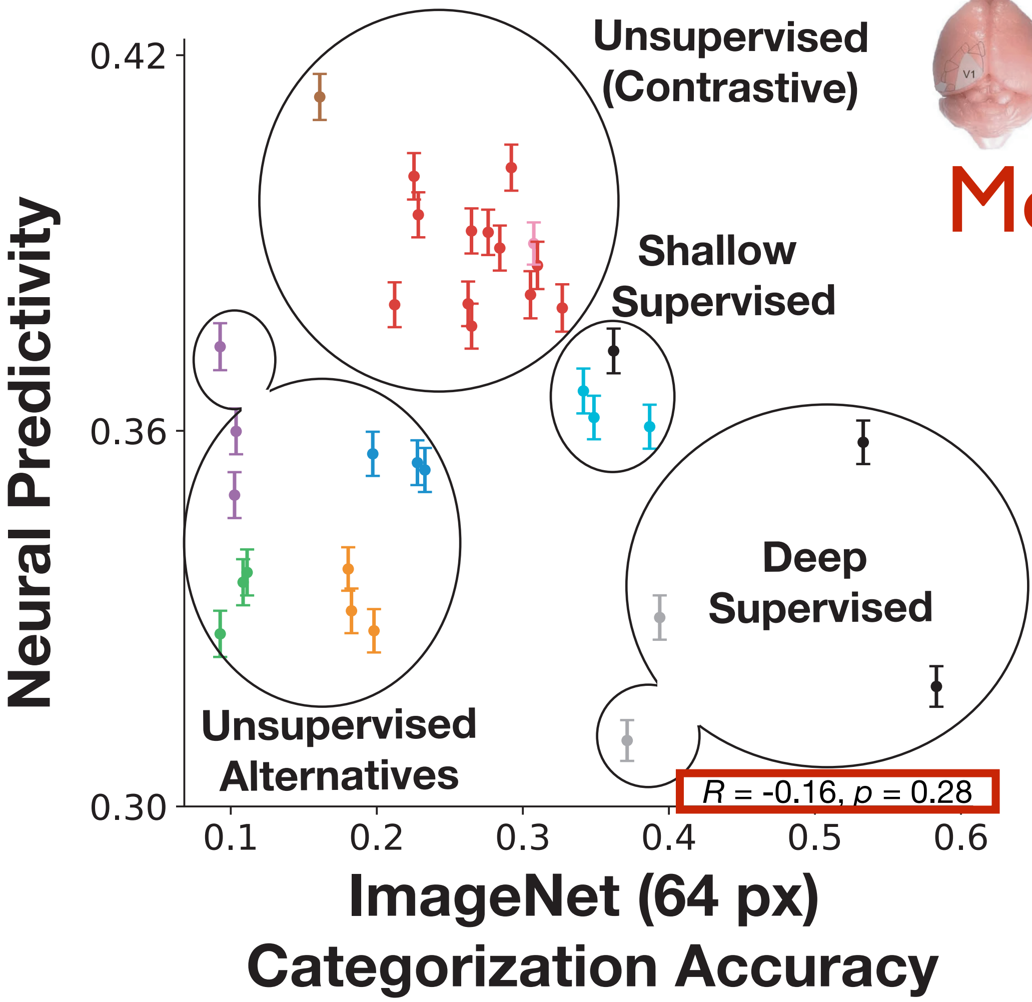
Contrastive objectives outperform "whisking" visual proxy loss

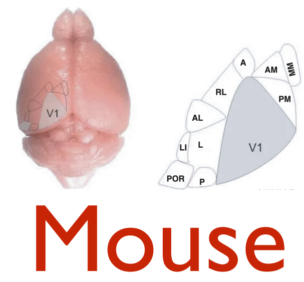
What is the ecological reason for unsupervised nets?

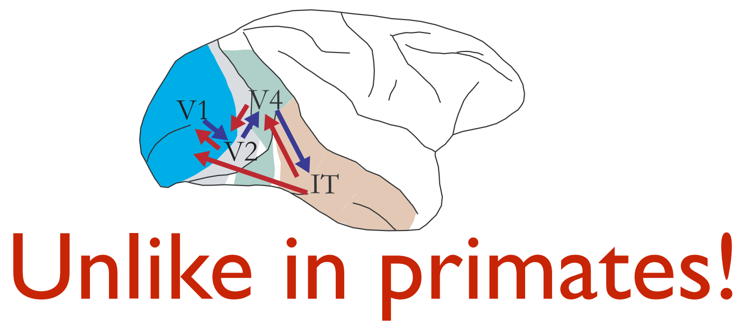
ImageNet categorization performance **NOT** correlated with neural predictivity



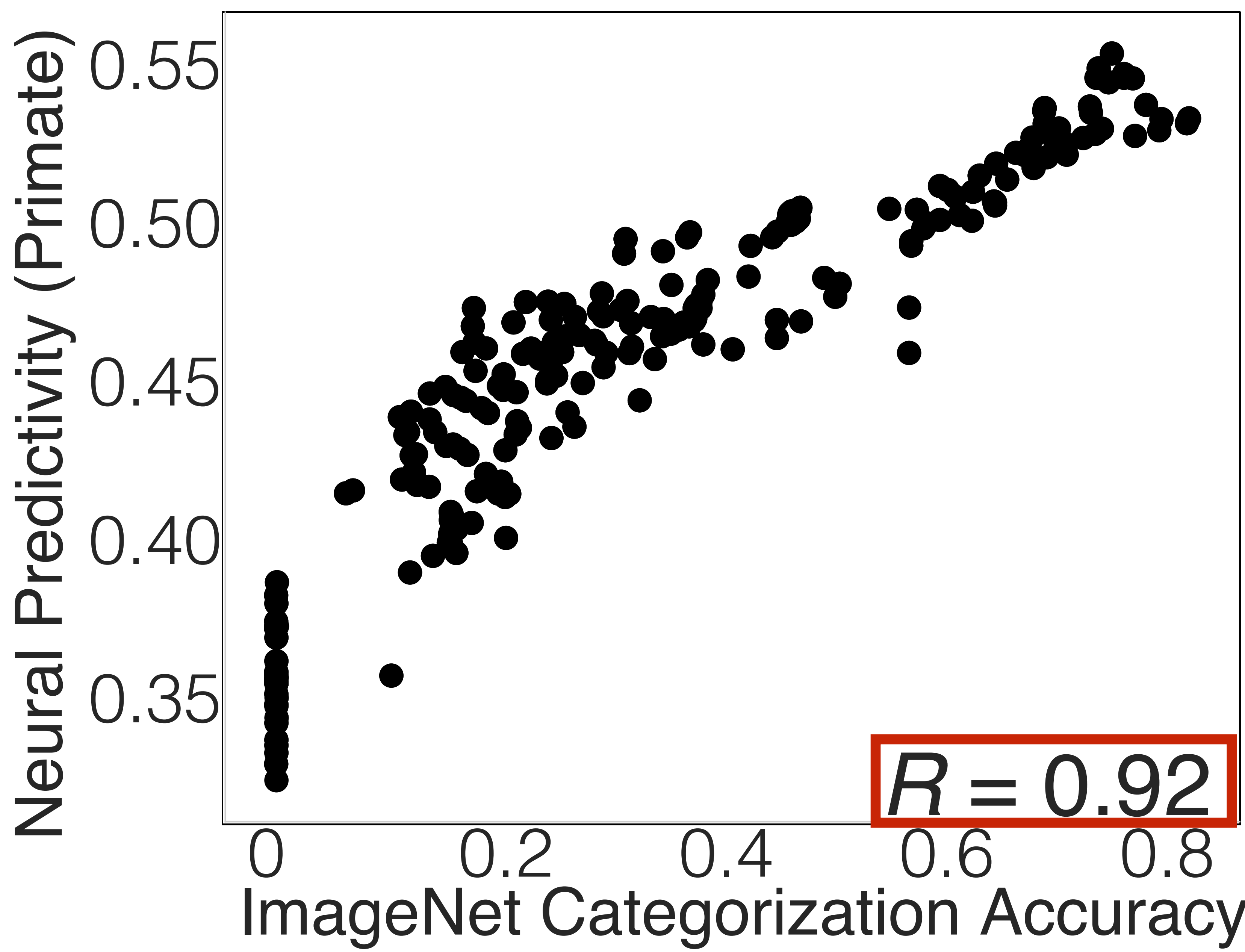
ImageNet categorization performance **NOT** correlated with neural predictivity







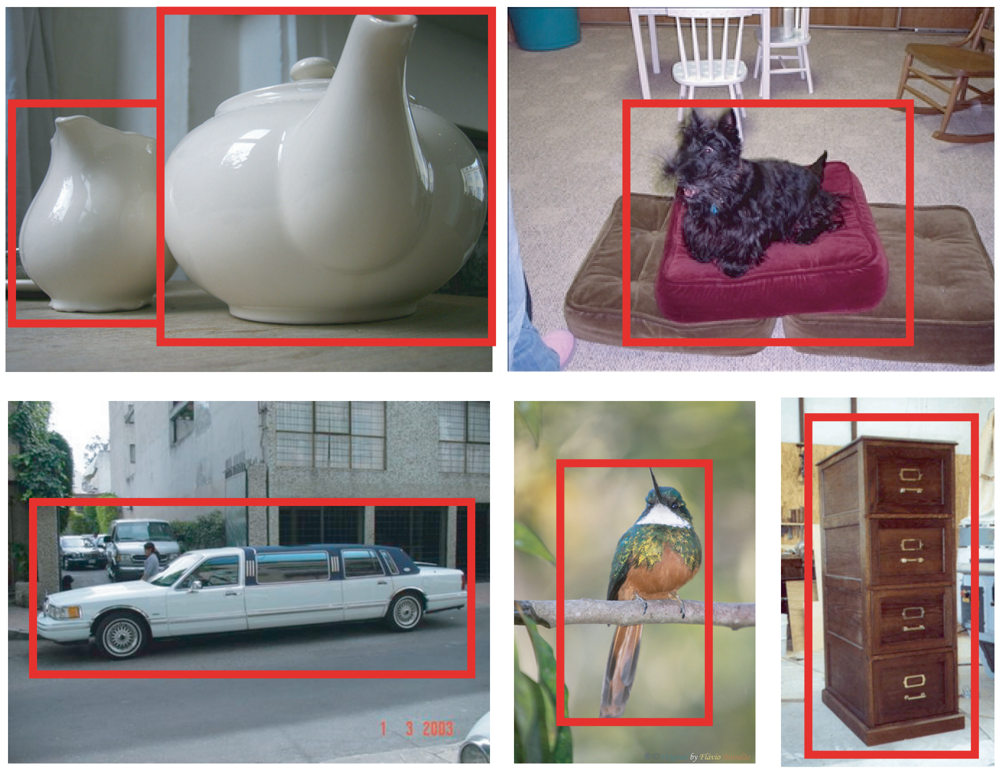
Schrimpf et al. 2018



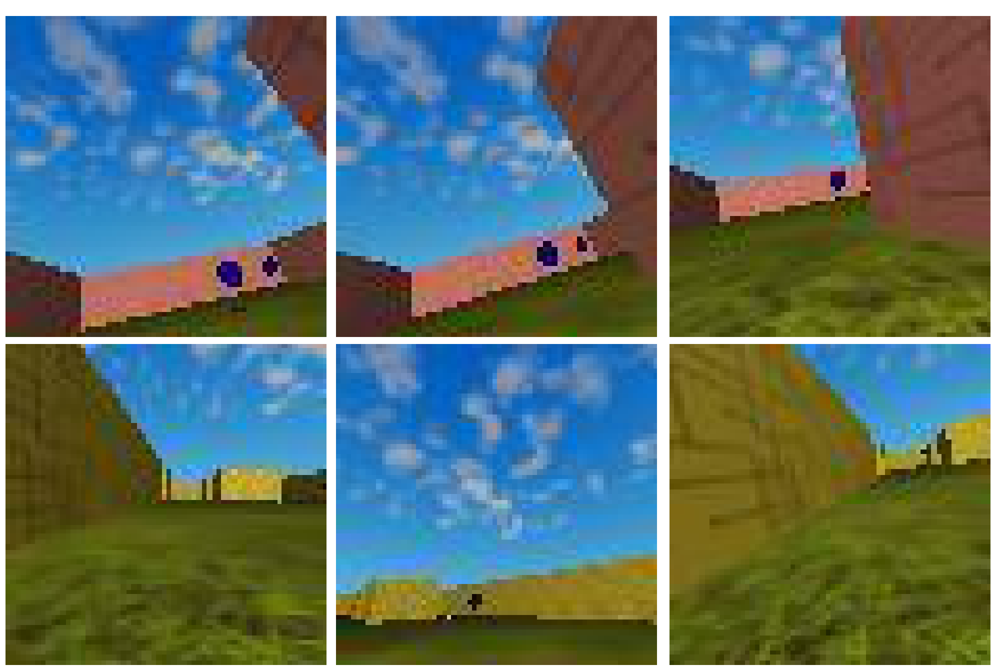
Assessing Task-Generality

Train

ImageNet

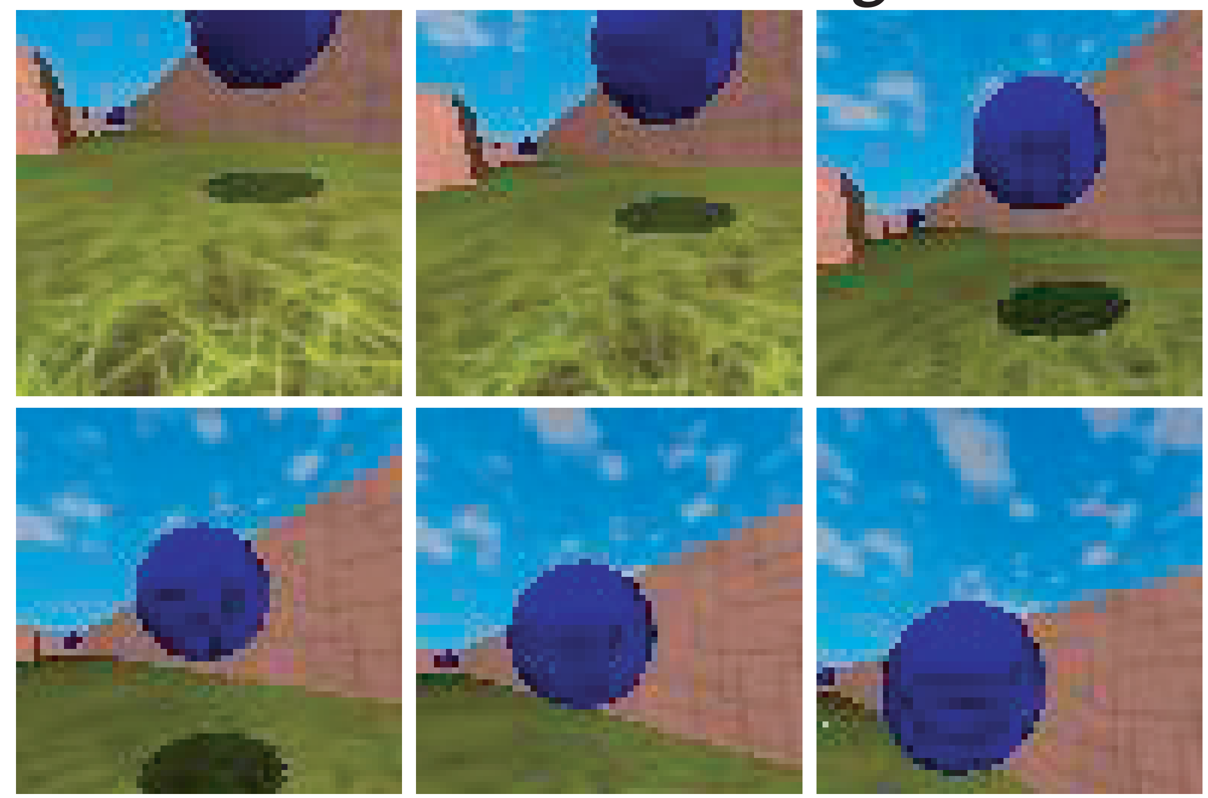


Maze Environment

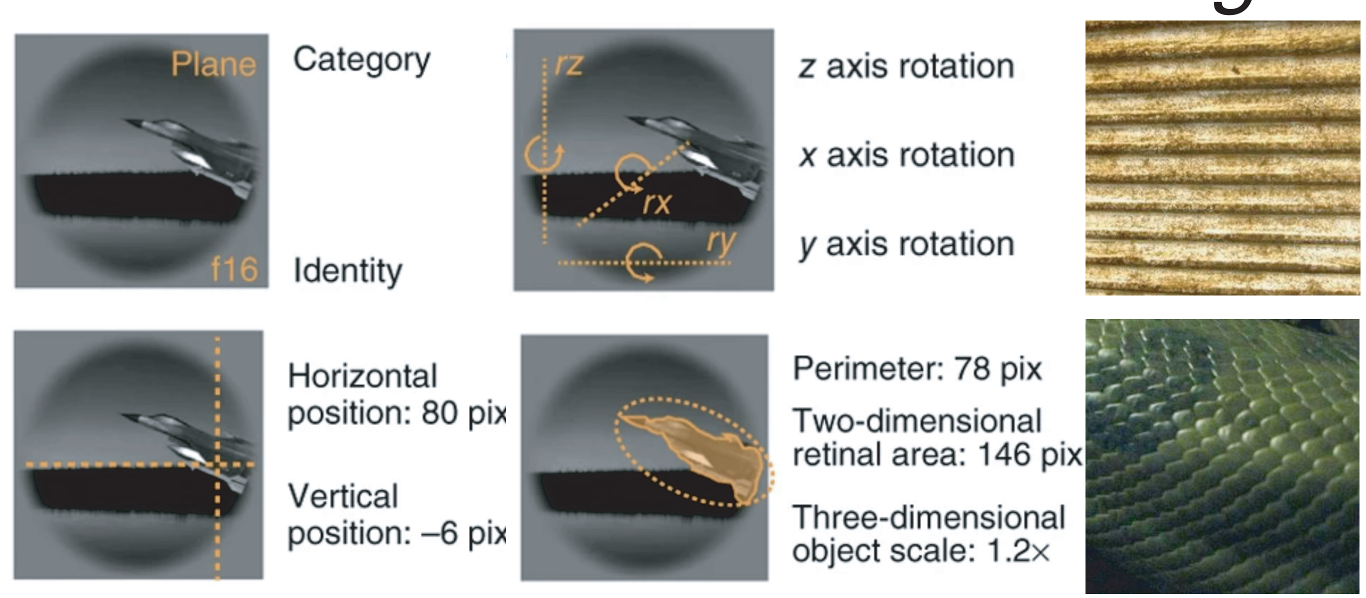


Evaluate

Reward-Based Navigation

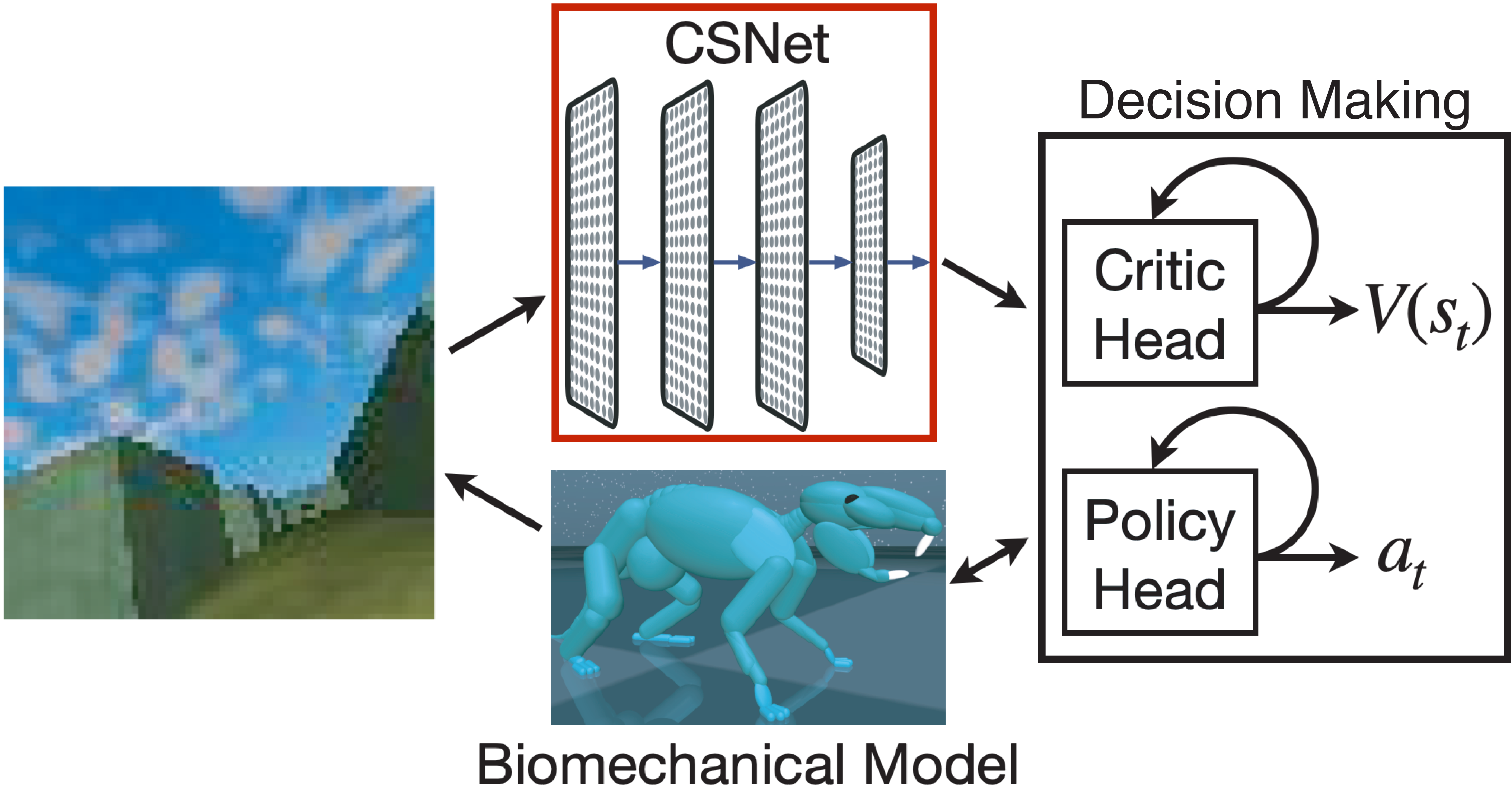


Visual Scene Understanding



Object properties *Texture*

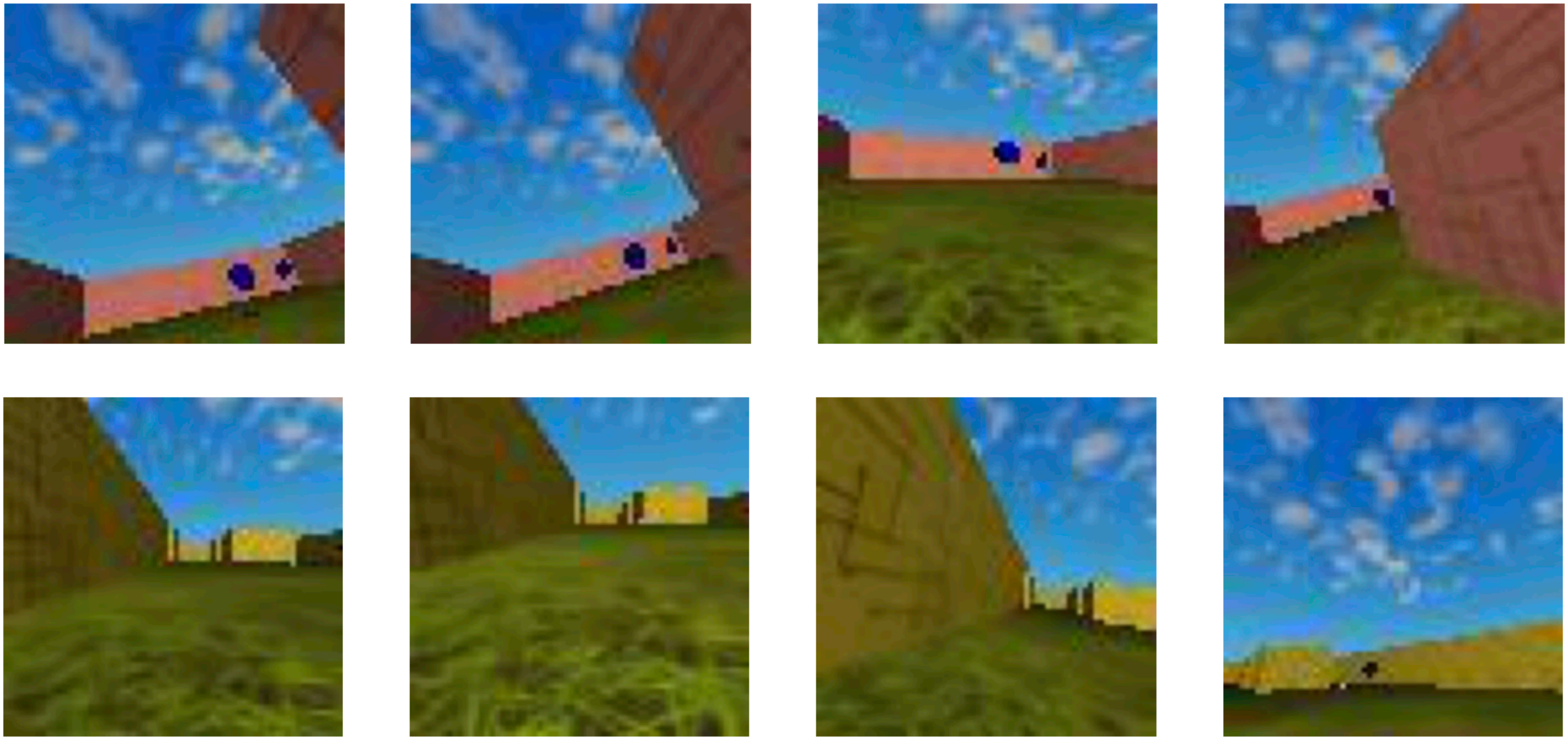
Schematic of Virtual Rodent



(*Merel*, Aldorando*, Marshal* et al. 2020*)

Navigation task in "Rodent Mazes" environment

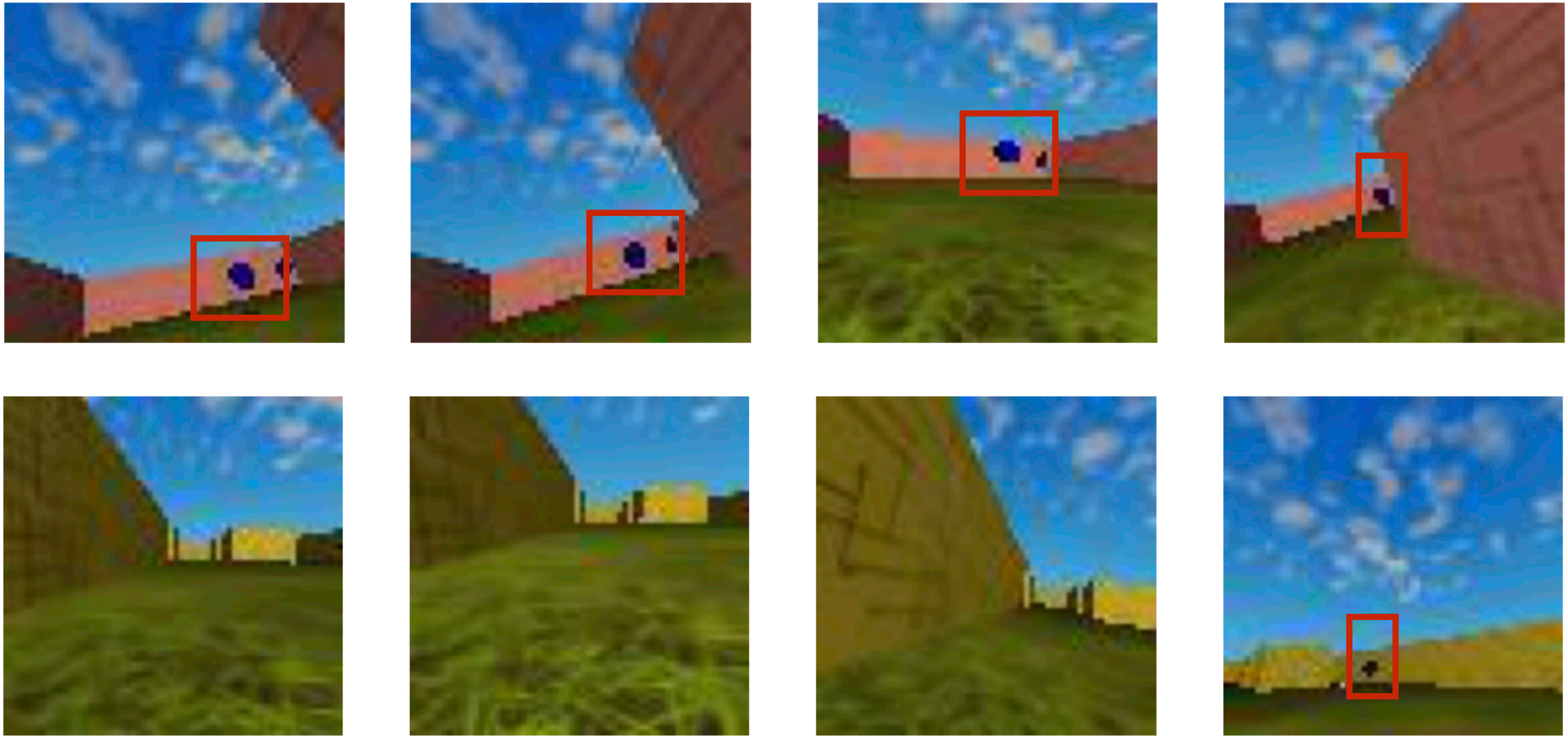
Merel, Aldorando*, Marshal* et al. 2020; Gulcehre et al. 2021*



Requires keeping track of history over long timescales with high-dimensional, continuous inputs

Navigation task in "Rodent Mazes" environment

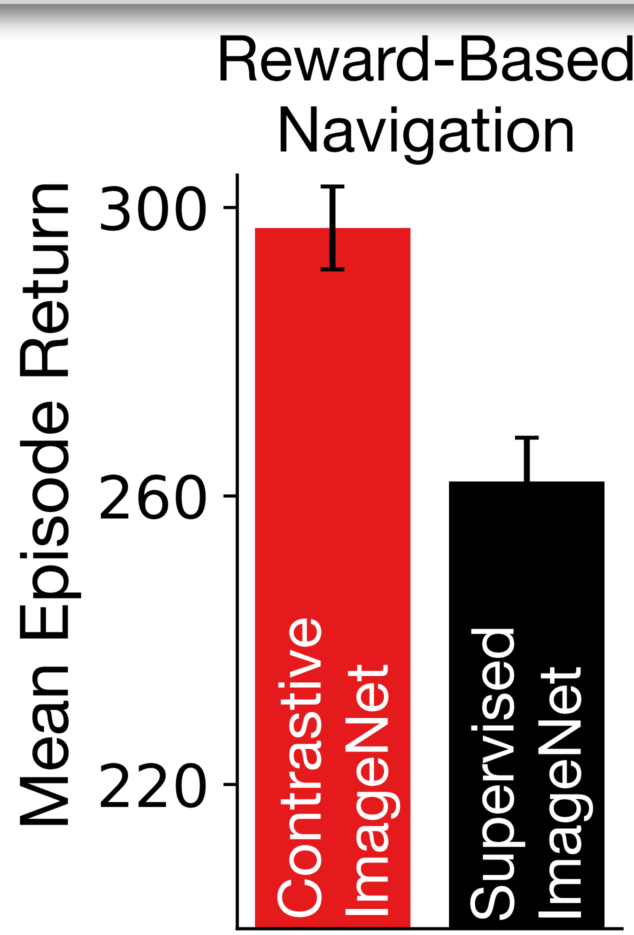
Merel, Aldorando*, Marshal* et al. 2020; Gulcehre et al. 2021*

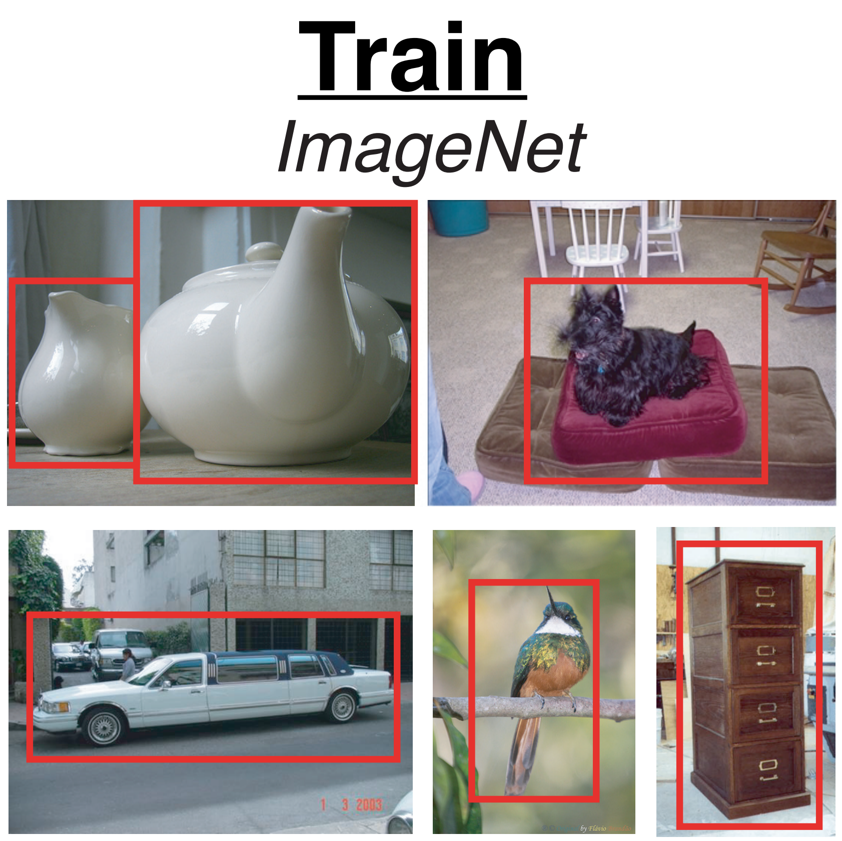


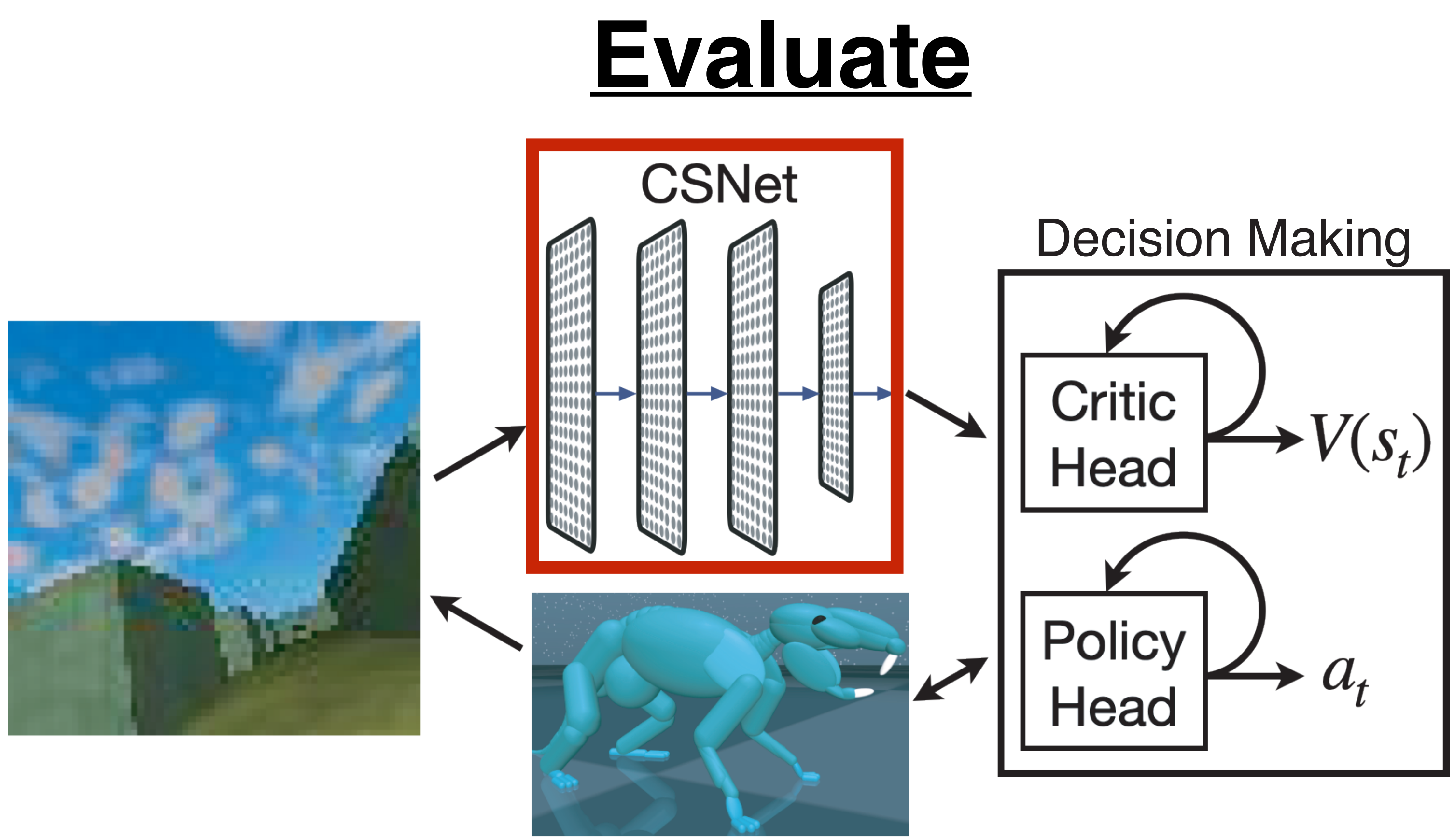
Requires keeping track of history over long timescales with high-dimensional, continuous inputs

Contrastive models yield better transfer performance

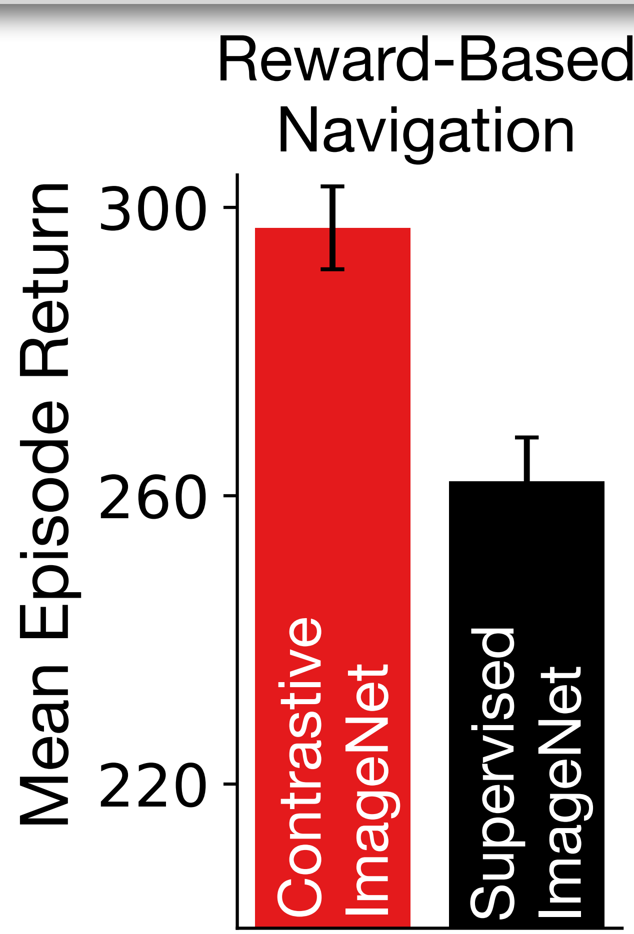
Contrastive models yield better transfer performance

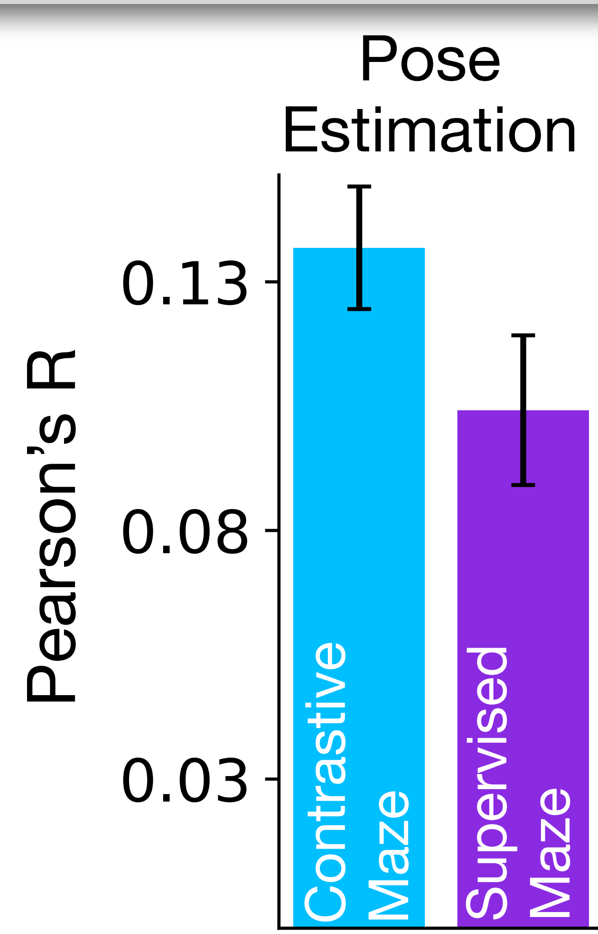






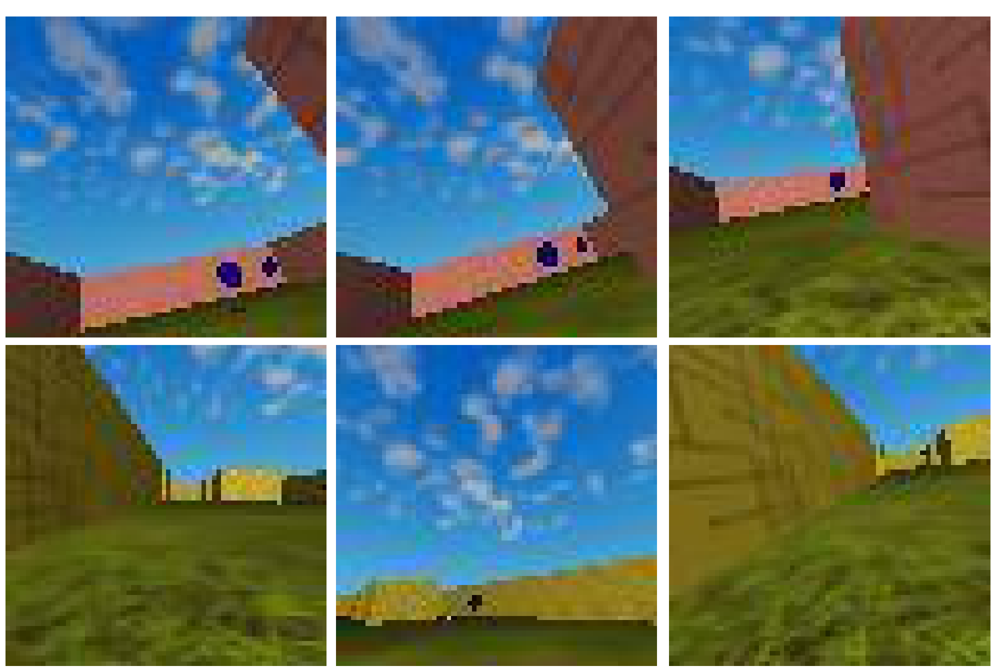
Contrastive models yield better transfer performance





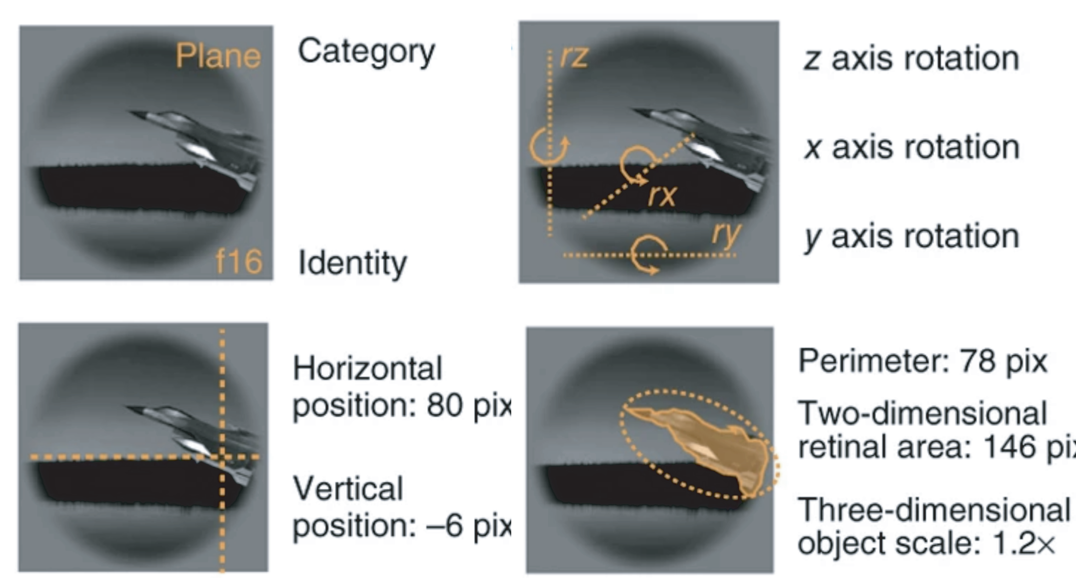
Train

Maze Environment

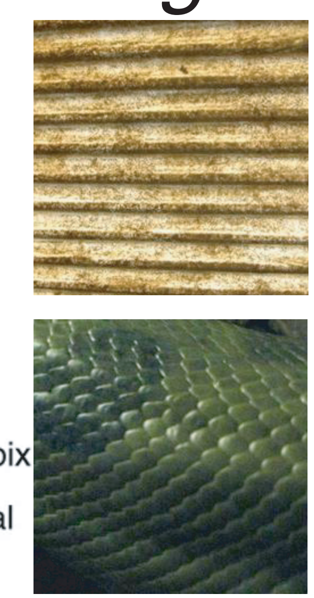


Evaluate

Visual Scene Understanding

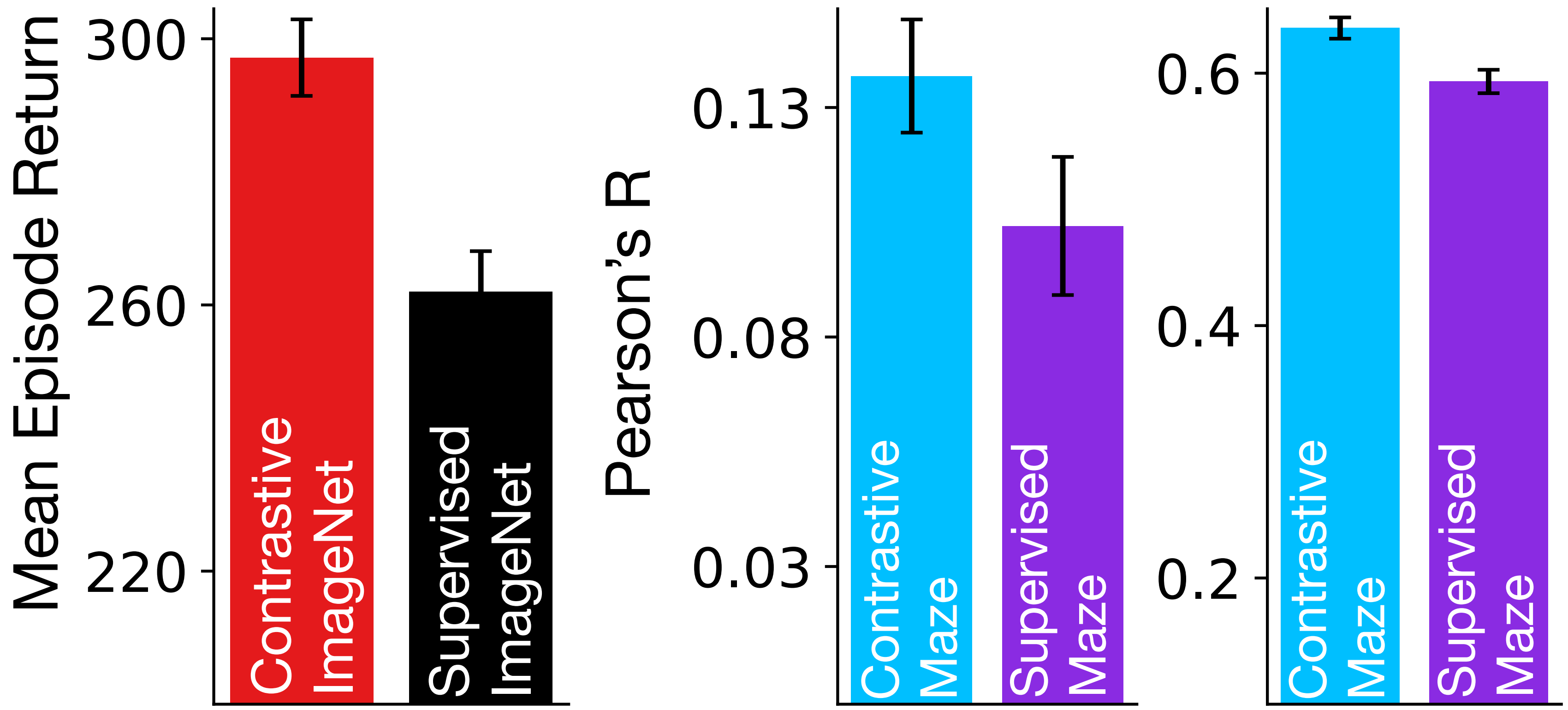


Object properties



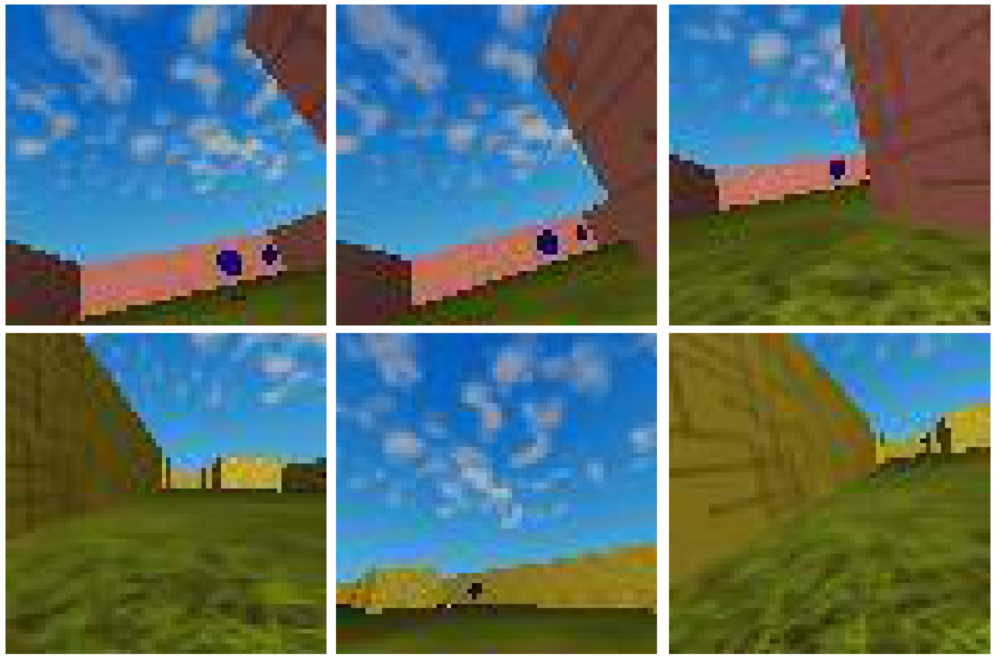
Texture

Contrastive models yield better transfer performance



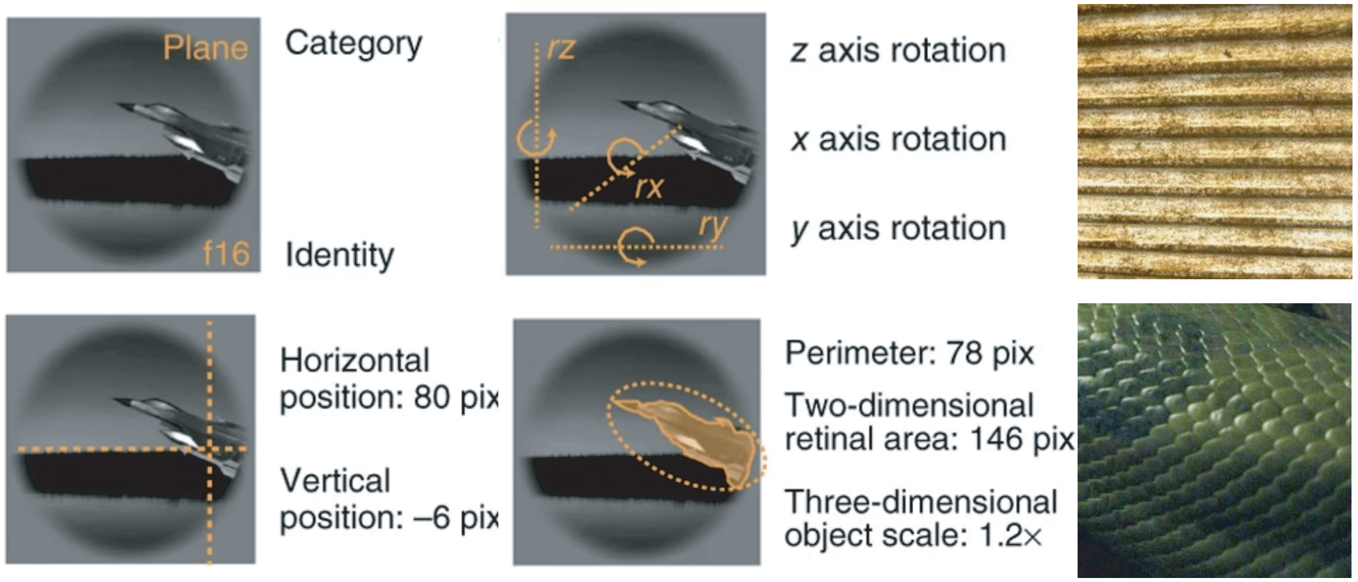
Train

Maze Environment



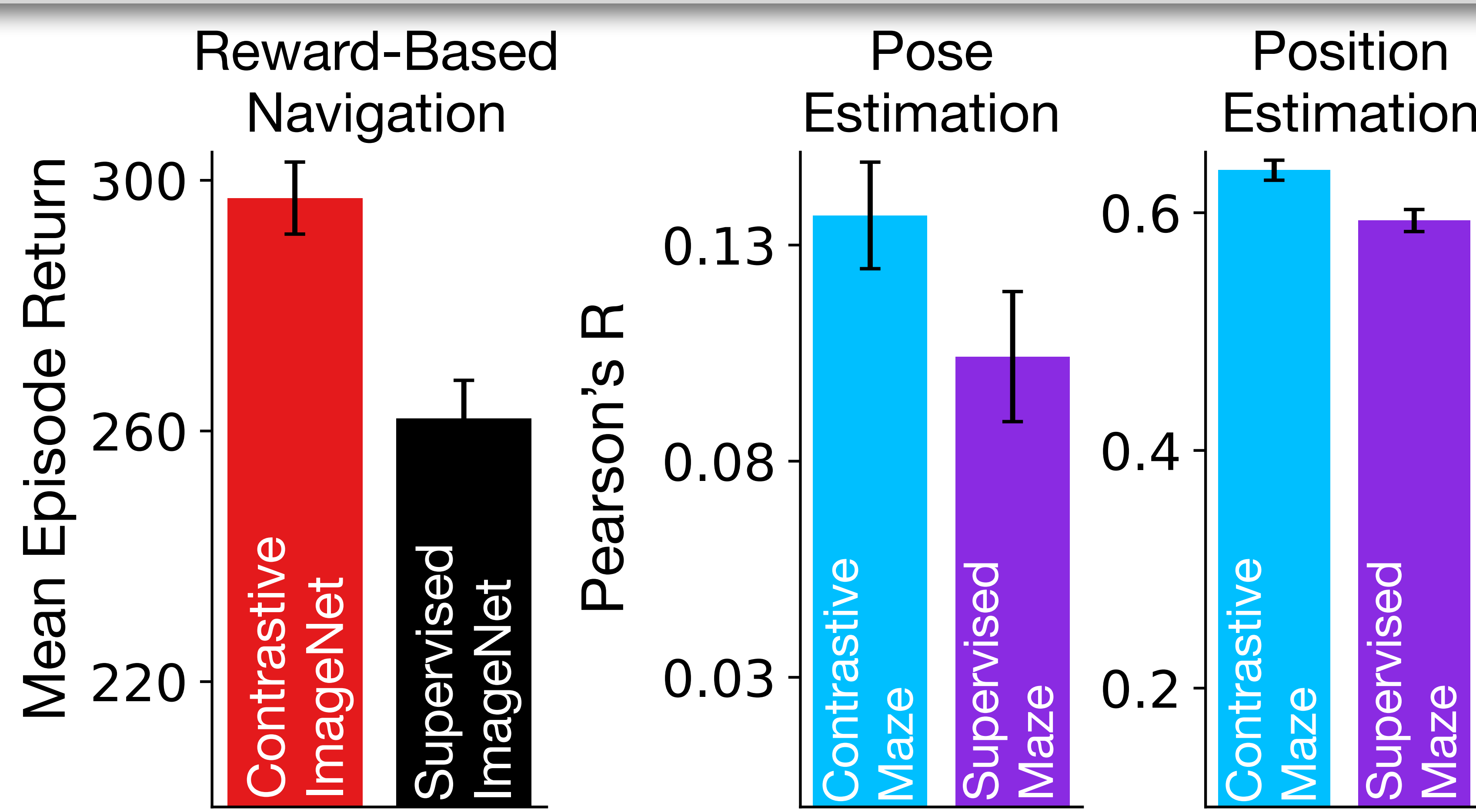
Evaluate

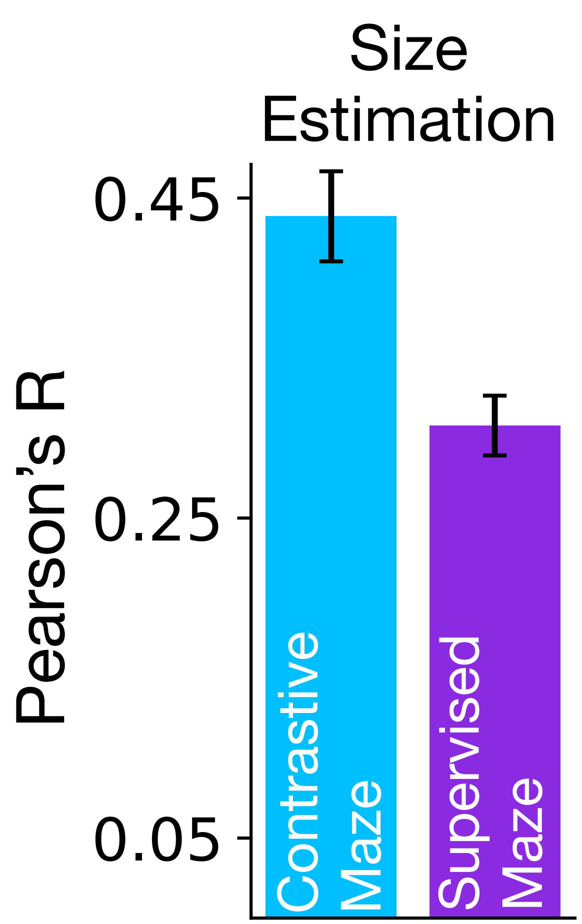
Visual Scene Understanding



Object properties *Texture*

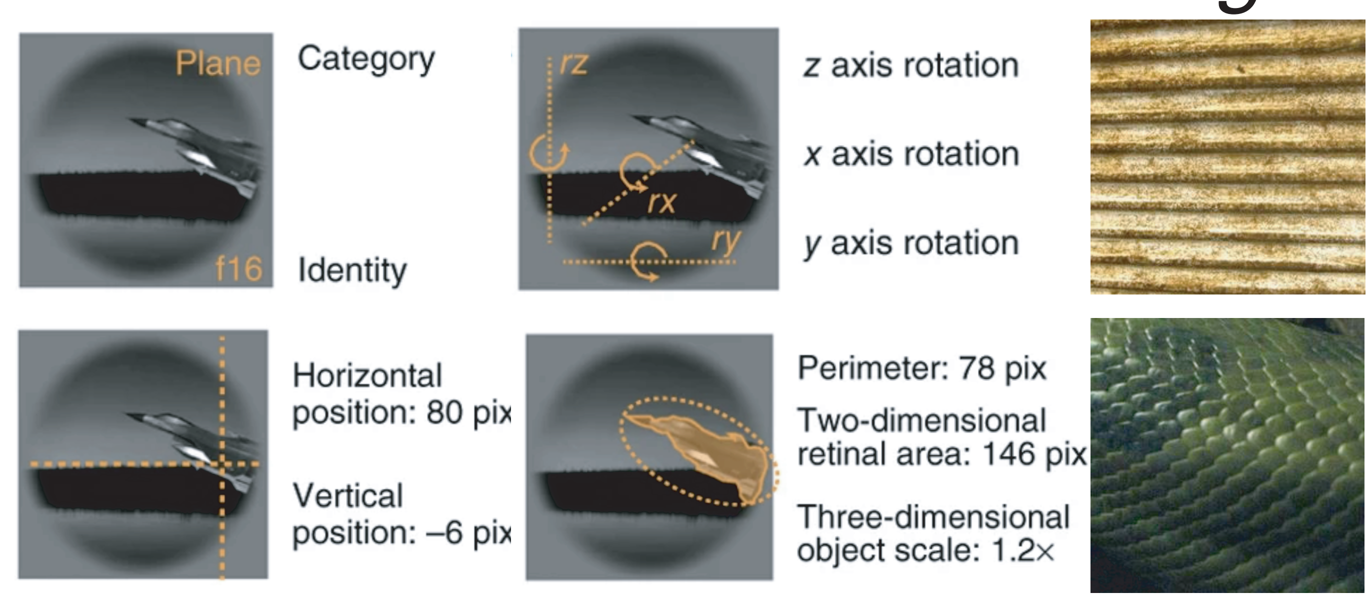
Contrastive models yield better transfer performance





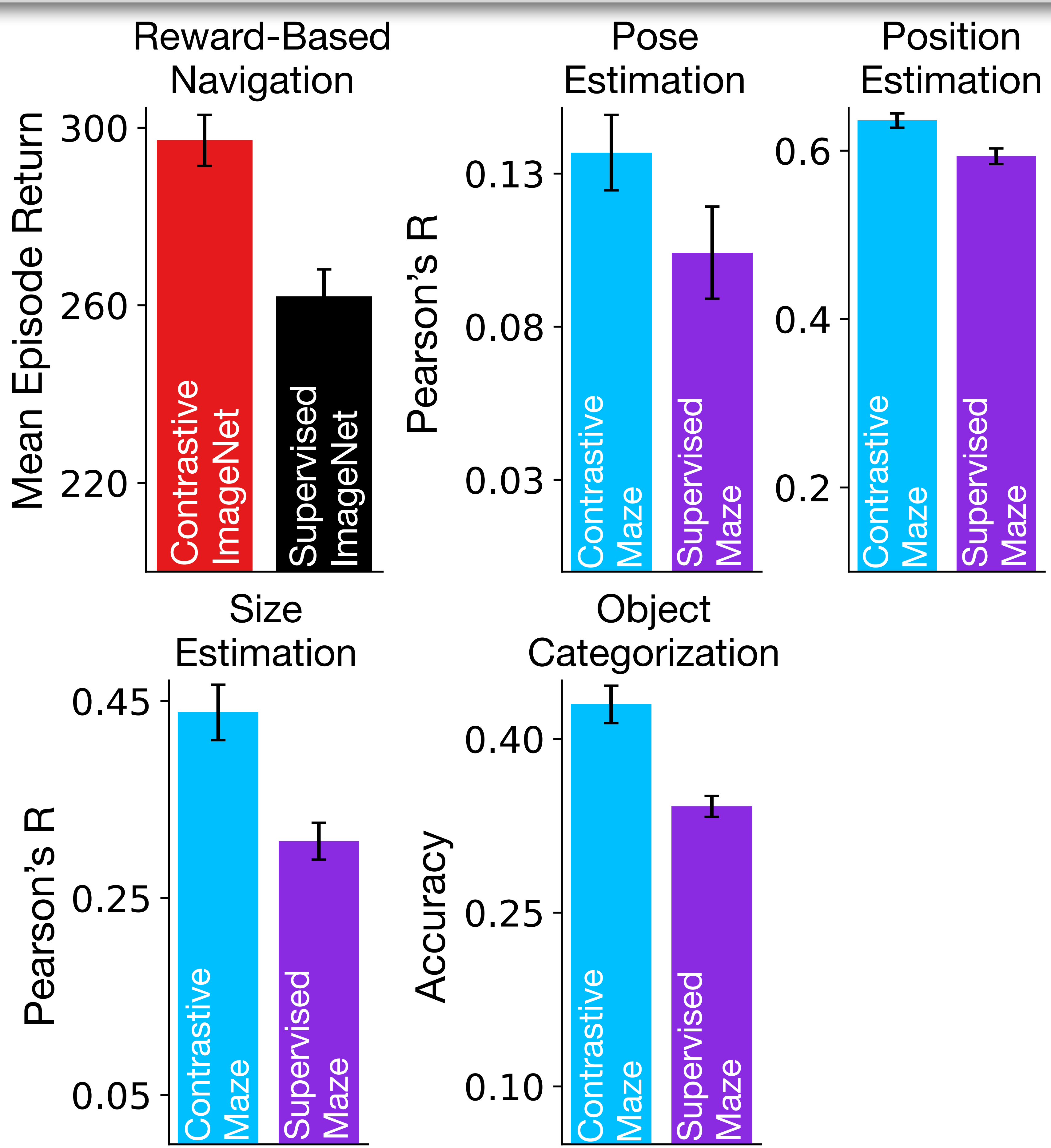
Evaluate

Visual Scene Understanding

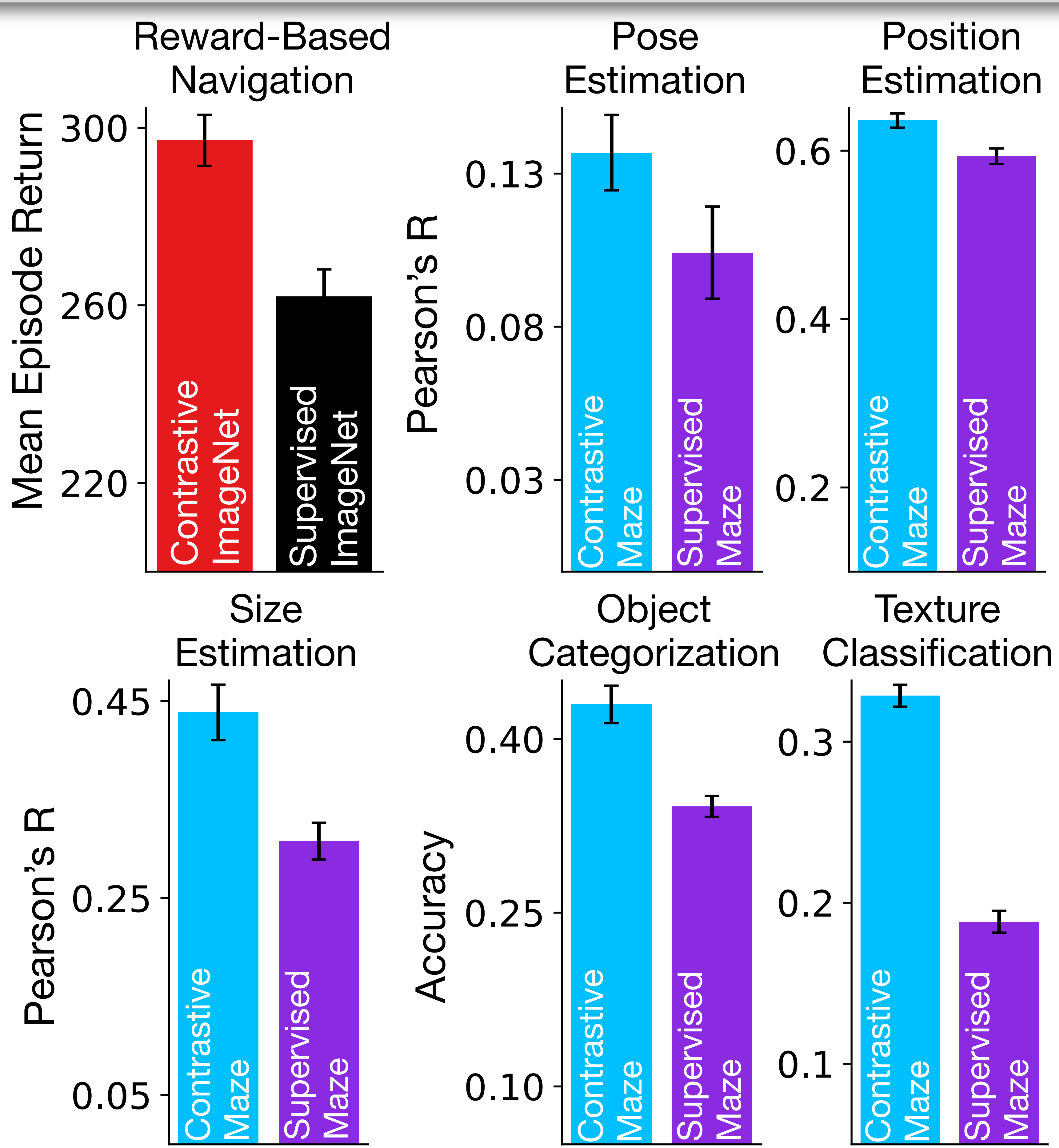


Object properties *Texture*

Contrastive models yield better transfer performance



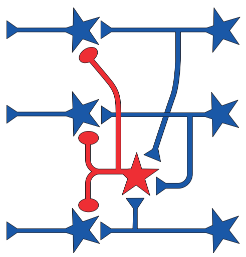
Contrastive models yield better transfer performance



Takeaways

A = *architecture class*

1. *"Circuit"*

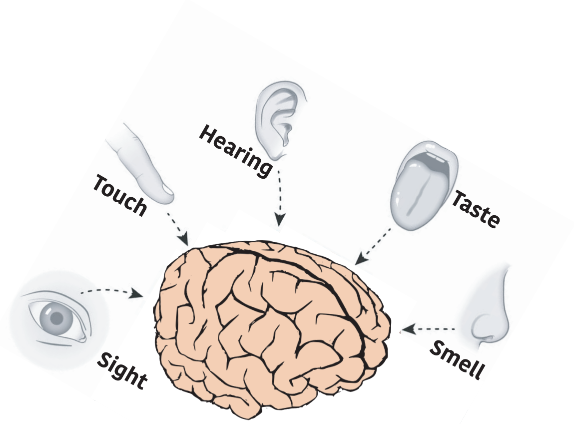


T = *task loss*

3. *"Ecological niche/behavior"*

<u>Neurobiological Puzzle:</u>

What are the core differences between mouse visual cortex and the primate ventral stream?



2. *"Environment"*

D = *data stream*

Takeaways

A = *architecture class*

1. *"Circuit"*

shallow

T = *task loss*

3. *"Ecological niche/behavior"*

unsupervised

Neurobiological Puzzle:

What are the core differences between mouse visual cortex and the primate ventral stream?

Partial Resolution:

Mouse visual cortex is a general-purpose system utilizing its limited resources to perform a variety of visual tasks.

In contrast to the deep, high-resolution, and *task-specific* primate ventral stream.

low resolution

2. *"Environment"*

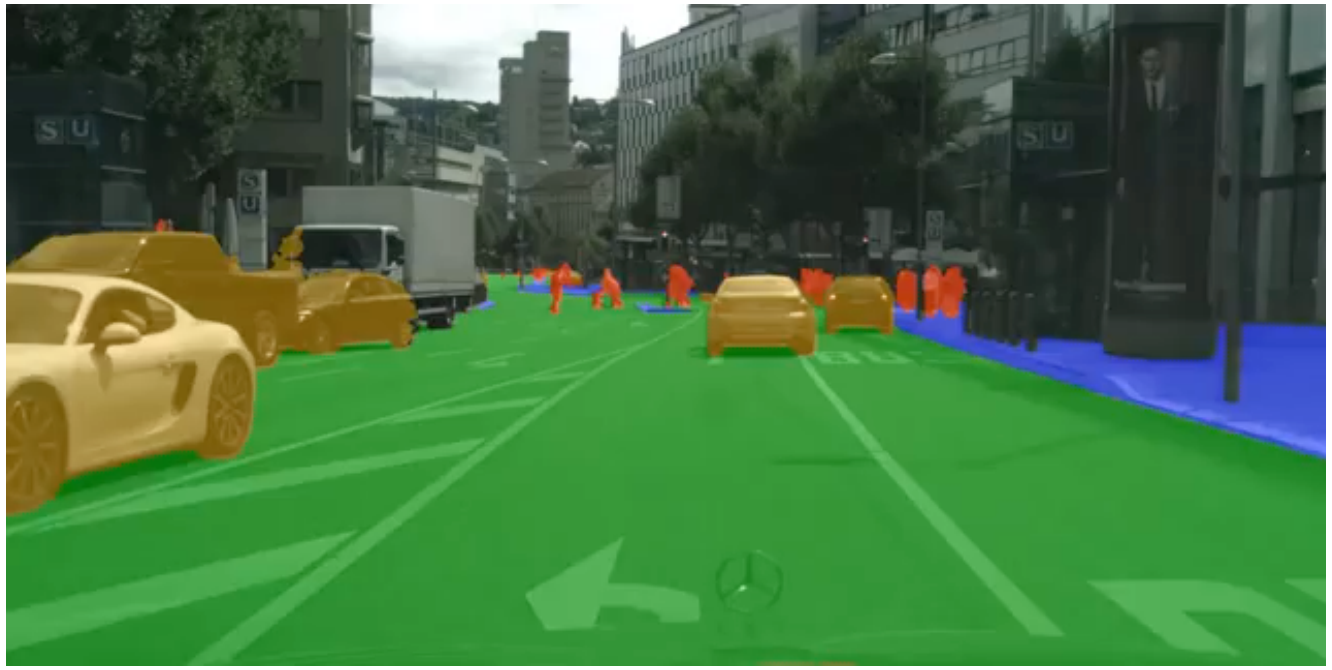
D = *data stream*

Outline

- Recurrent Connections in the Primate Ventral Stream
- Goal-Driven Models of Mouse Visual Cortex
- Heterogeneity in Rodent Medial Entorhinal Cortex
- Future Directions

From Neurons to Behavior

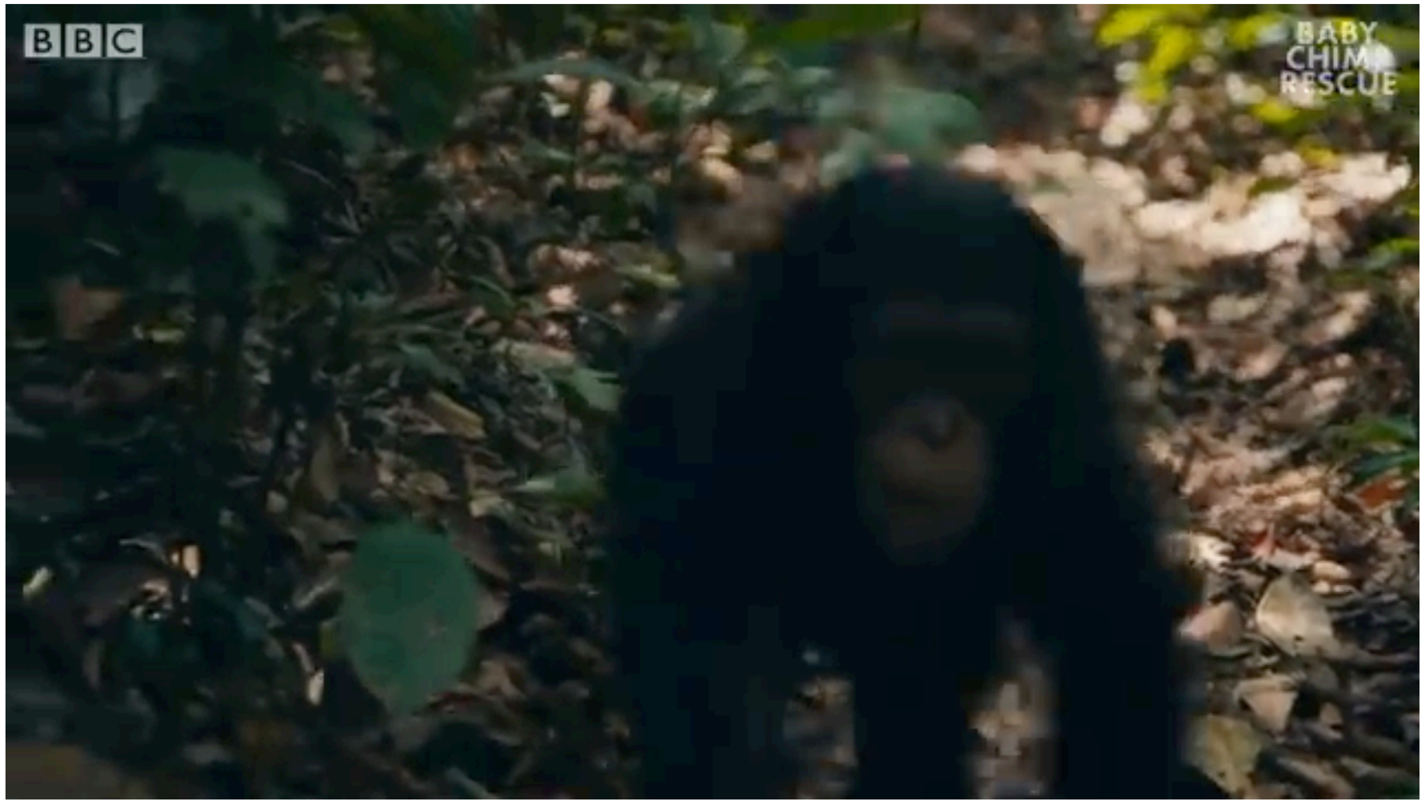
Scene Understanding



Multi-Step Planning

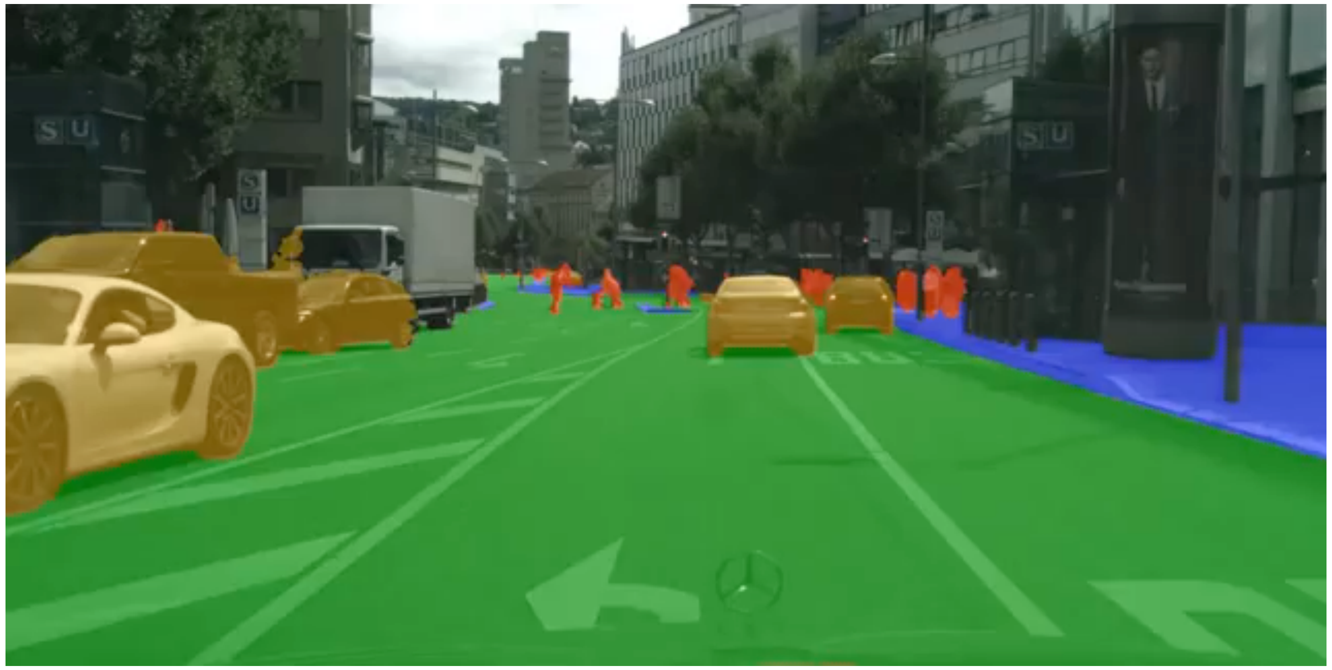
Navigation

Flexible Embodiment



From Neurons to Behavior

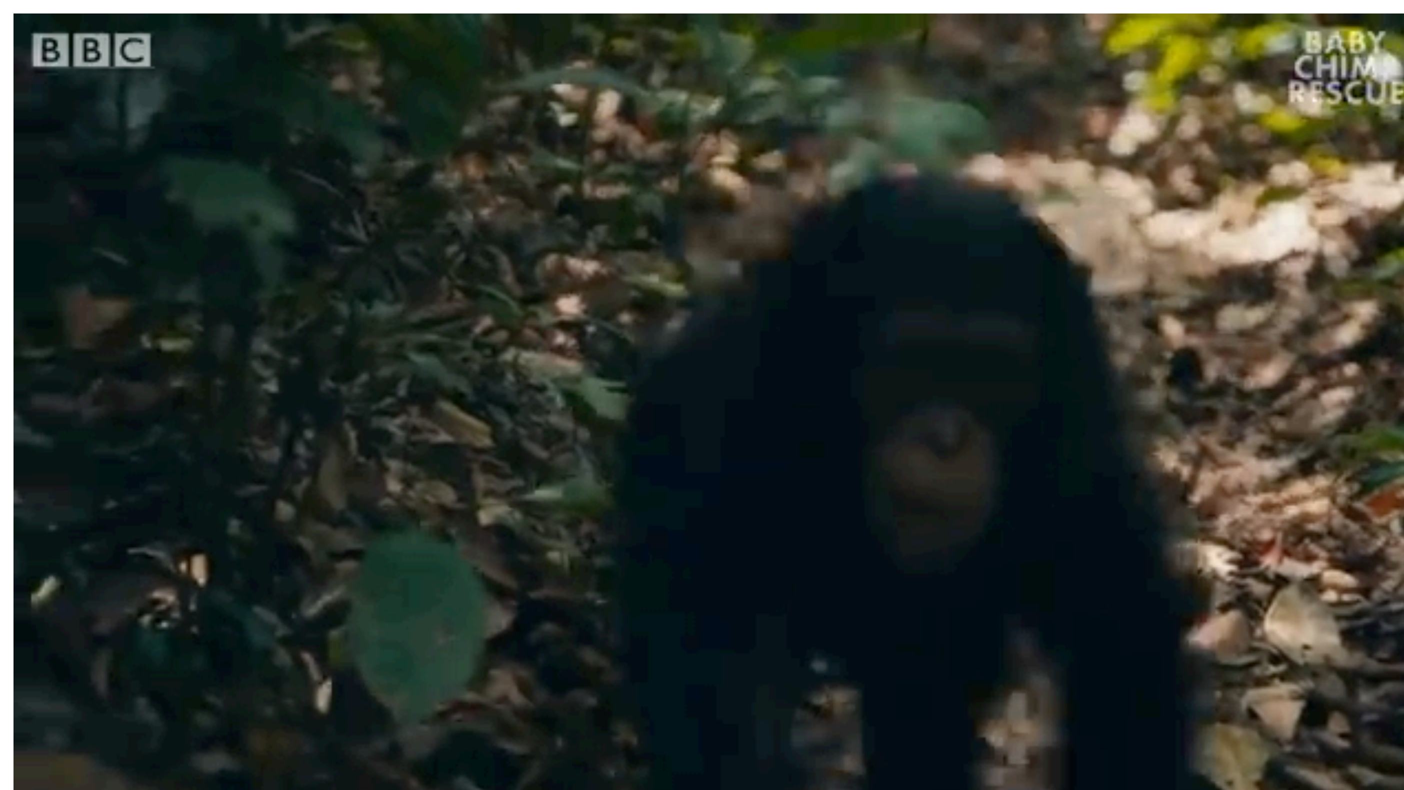
Scene Understanding



Multi-Step Planning

Navigation

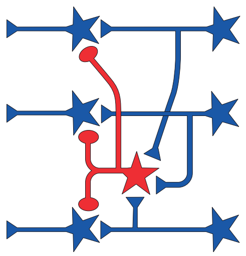
Flexible Embodiment



Heterogeneity in Rodent Medial Entorhinal Cortex

A = *architecture class*

1. *"Circuit"*



T = *task loss*

3. *"Ecological niche/behavior"*

A. Nayebi, *et al.*
Explaining Heterogeneity in Medial Entorhinal Cortex with Task-Driven Neural Networks.
NeurIPS 2021 (spotlight)

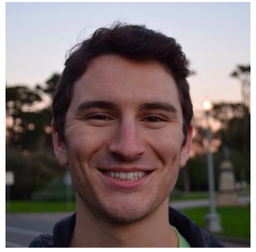

Alex Attinger


Malcolm Campbell


Kiah Hardcastle

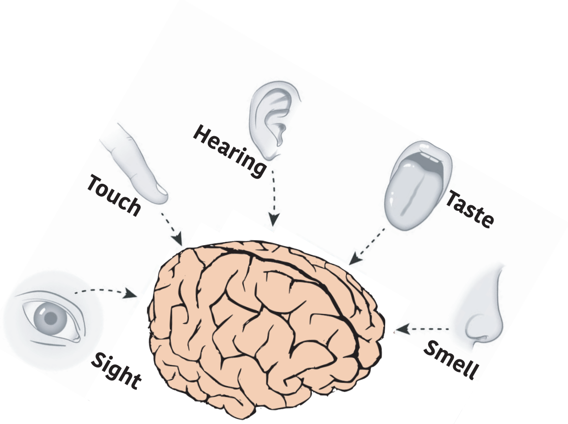

Isabel Low


Caitlin Mallory


Gabriel Mel


Ben Sorscher


Alex Williams



2. *"Environment"*

D = *data stream*


Surya Ganguli


Lisa Giocomo

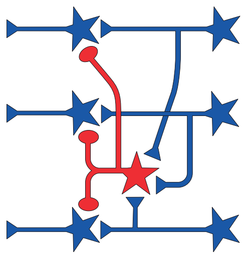

Daniel Yamins

Heterogeneity in Rodent Medial Entorhinal Cortex

A = *architecture class*

T = *task loss*

1. *"Circuit"*



3. *"Ecological niche/behavior"*

A. Nayebi, *et al.*
Explaining Heterogeneity in Medial Entorhinal Cortex with Task-Driven Neural Networks.
NeurIPS 2021 (spotlight)

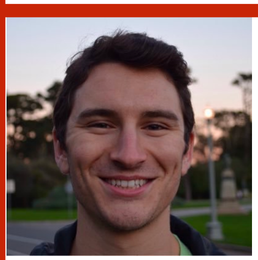

Alex Attinger


Malcolm Campbell


Kiah Hardcastle

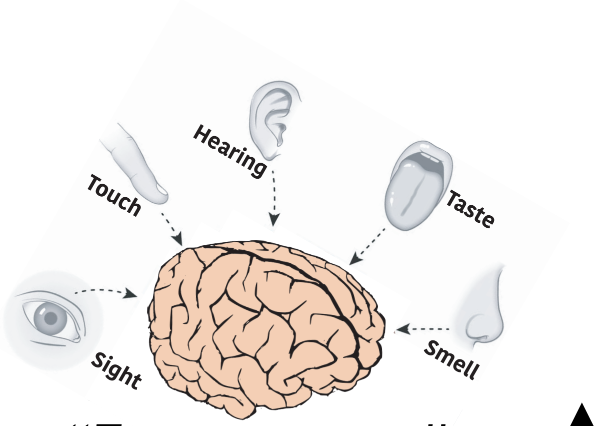

Isabel Low


Caitlin Mallory


Gabriel Mel


Ben Sorscher


Alex Williams



2. *"Environment"*

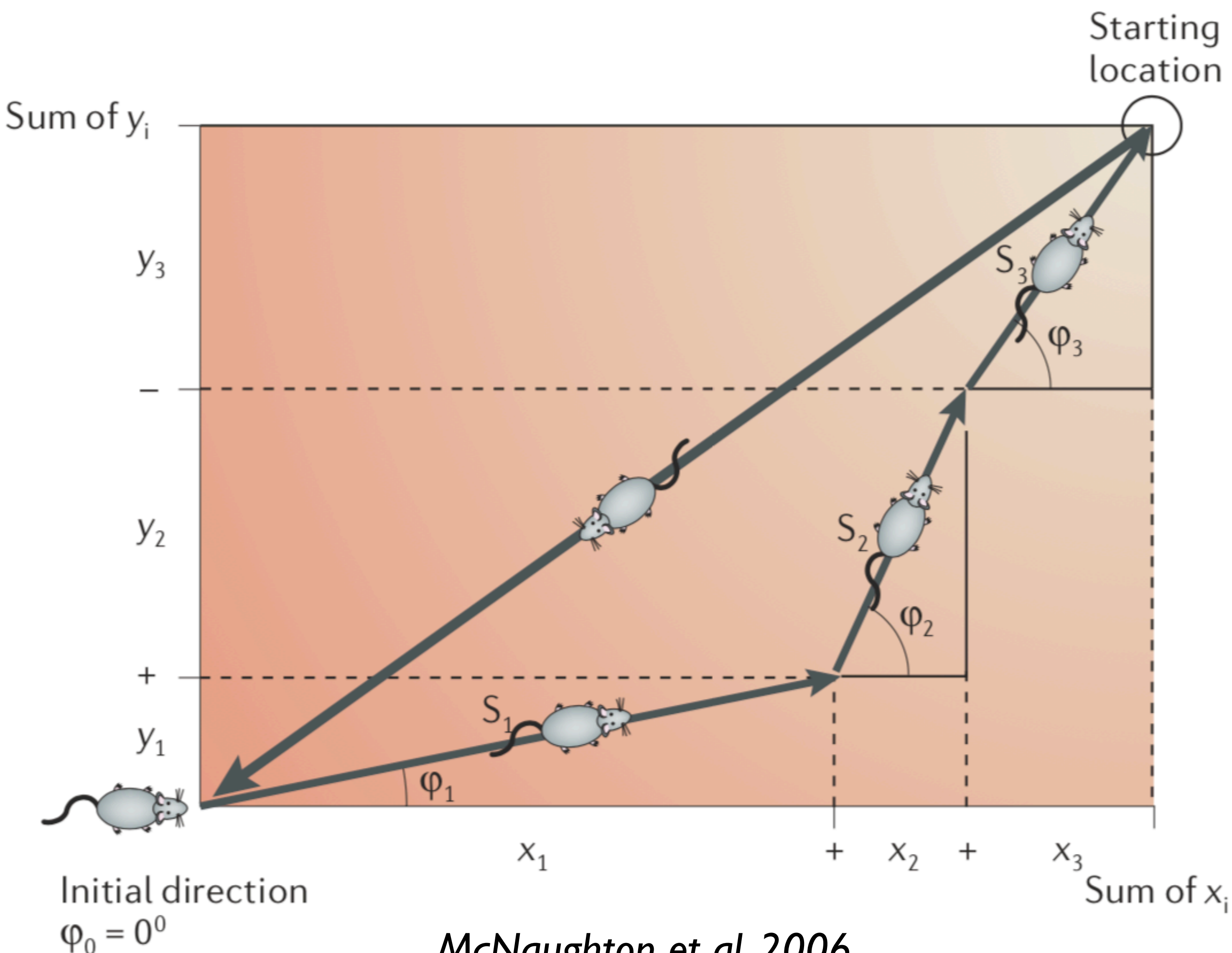
D = *data stream*


Surya Ganguli


Lisa Giocomo

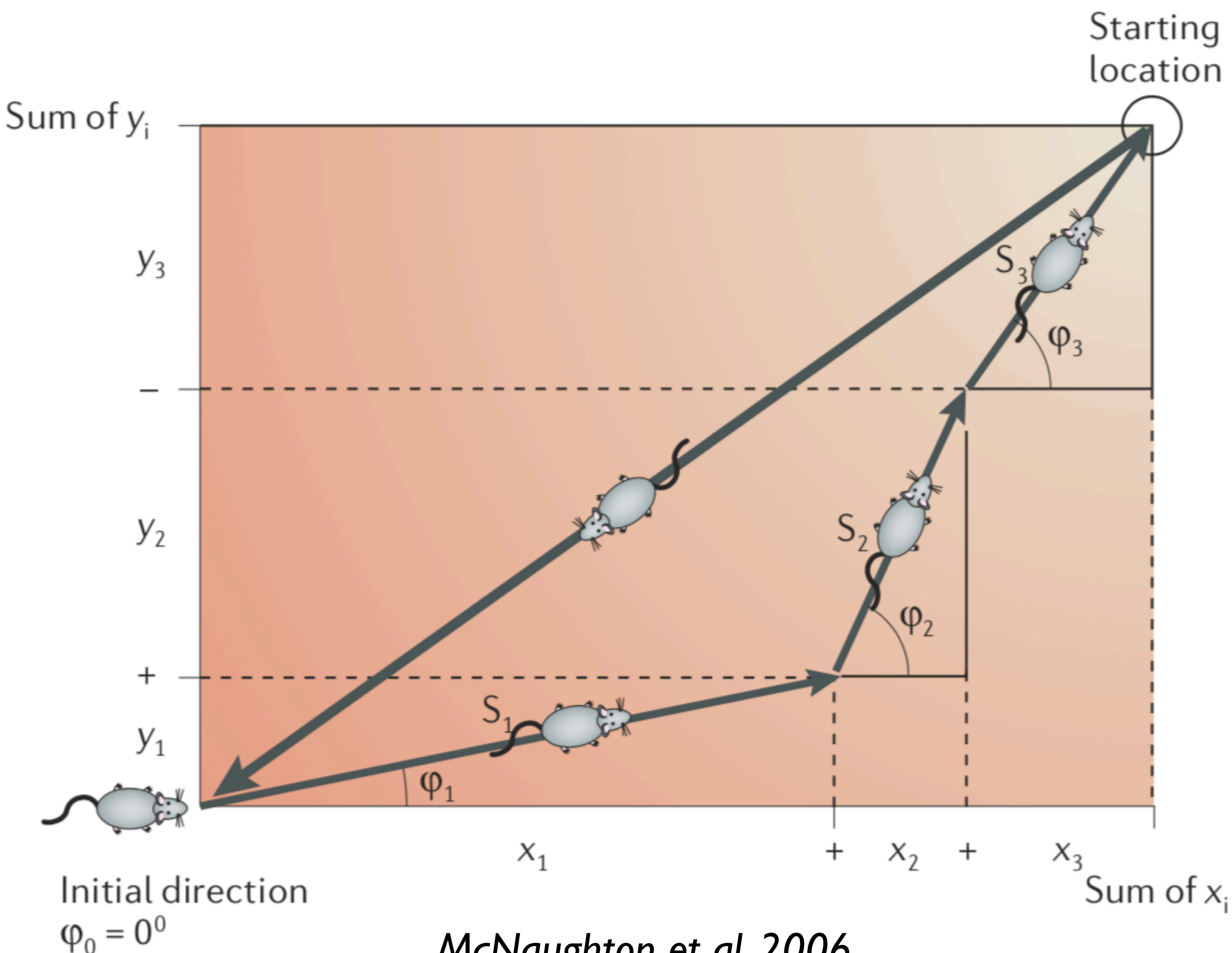

Daniel Yamins

Hippocampal-Entorhinal Spatial Map



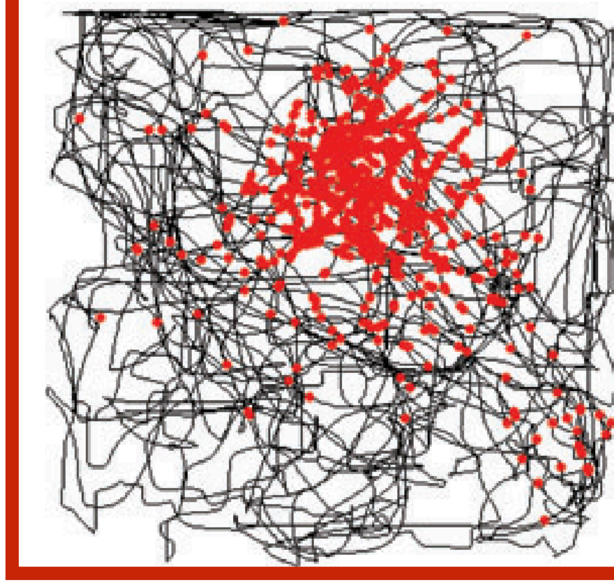
McNaughton et al. 2006

Hippocampal-Entorhinal Spatial Map

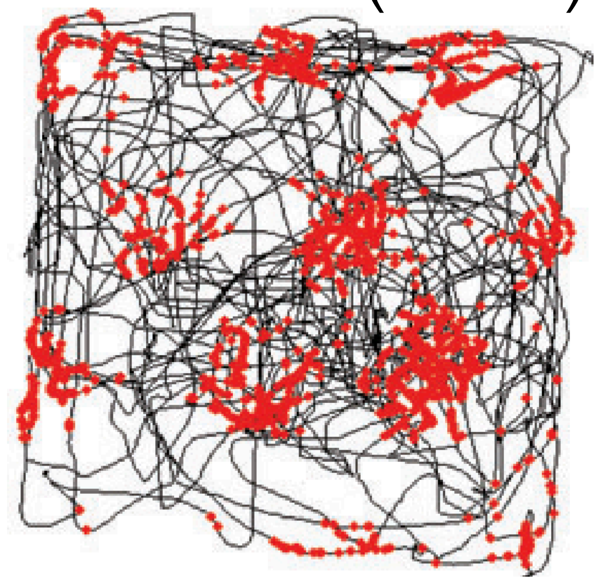


McNaughton et al. 2006

Place Cell (HPC)

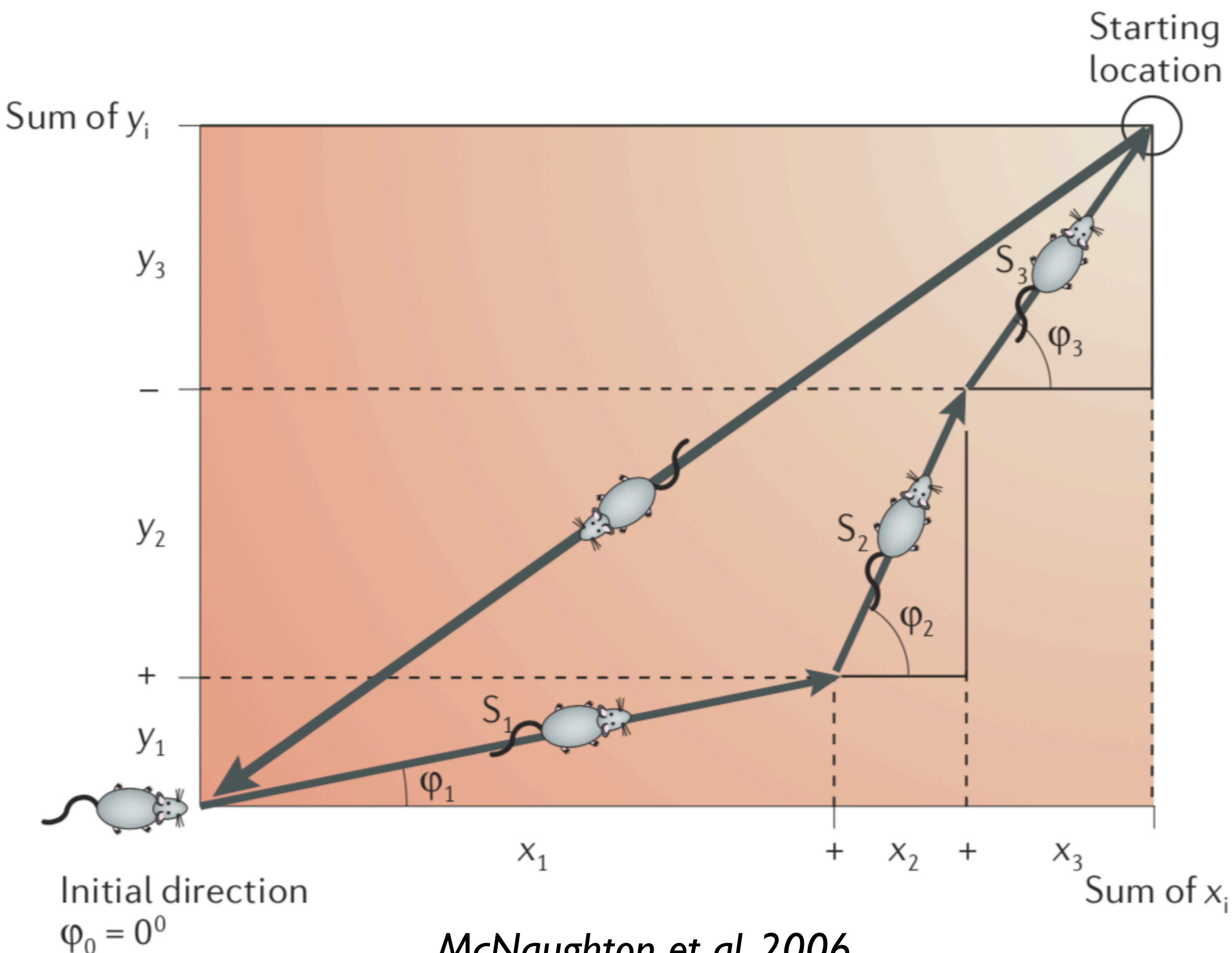


Grid Cell (MEC)



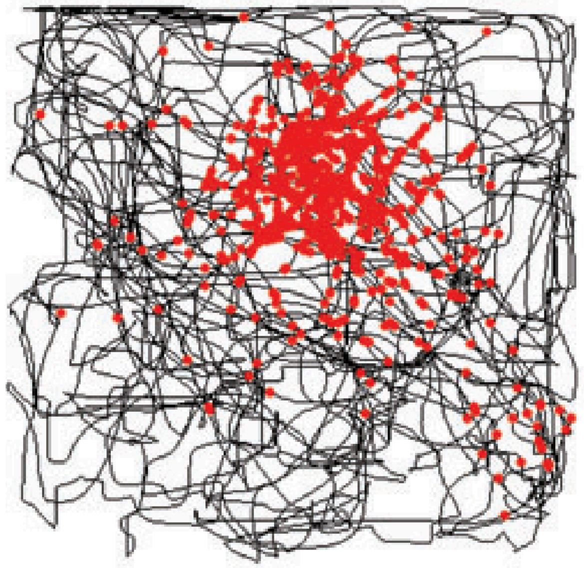
Moser et al. 2008

Hippocampal-Entorhinal Spatial Map

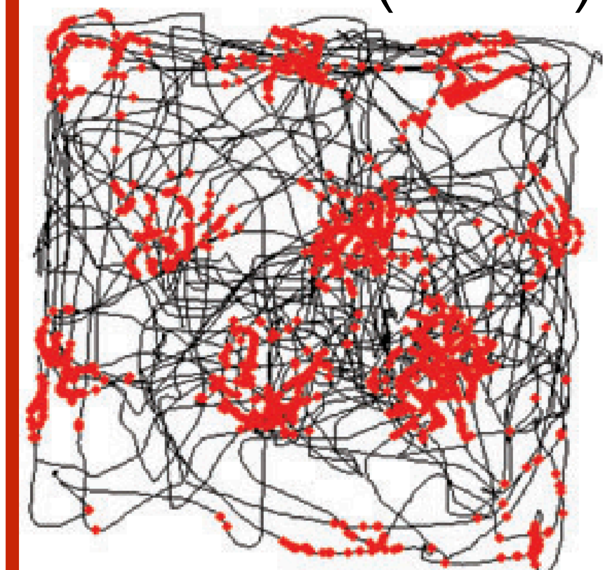


McNaughton et al. 2006

Place Cell (HPC)



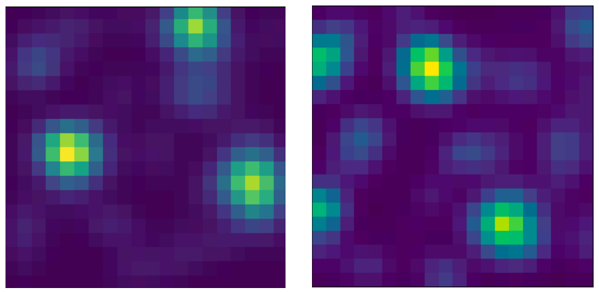
Grid Cell (MEC)



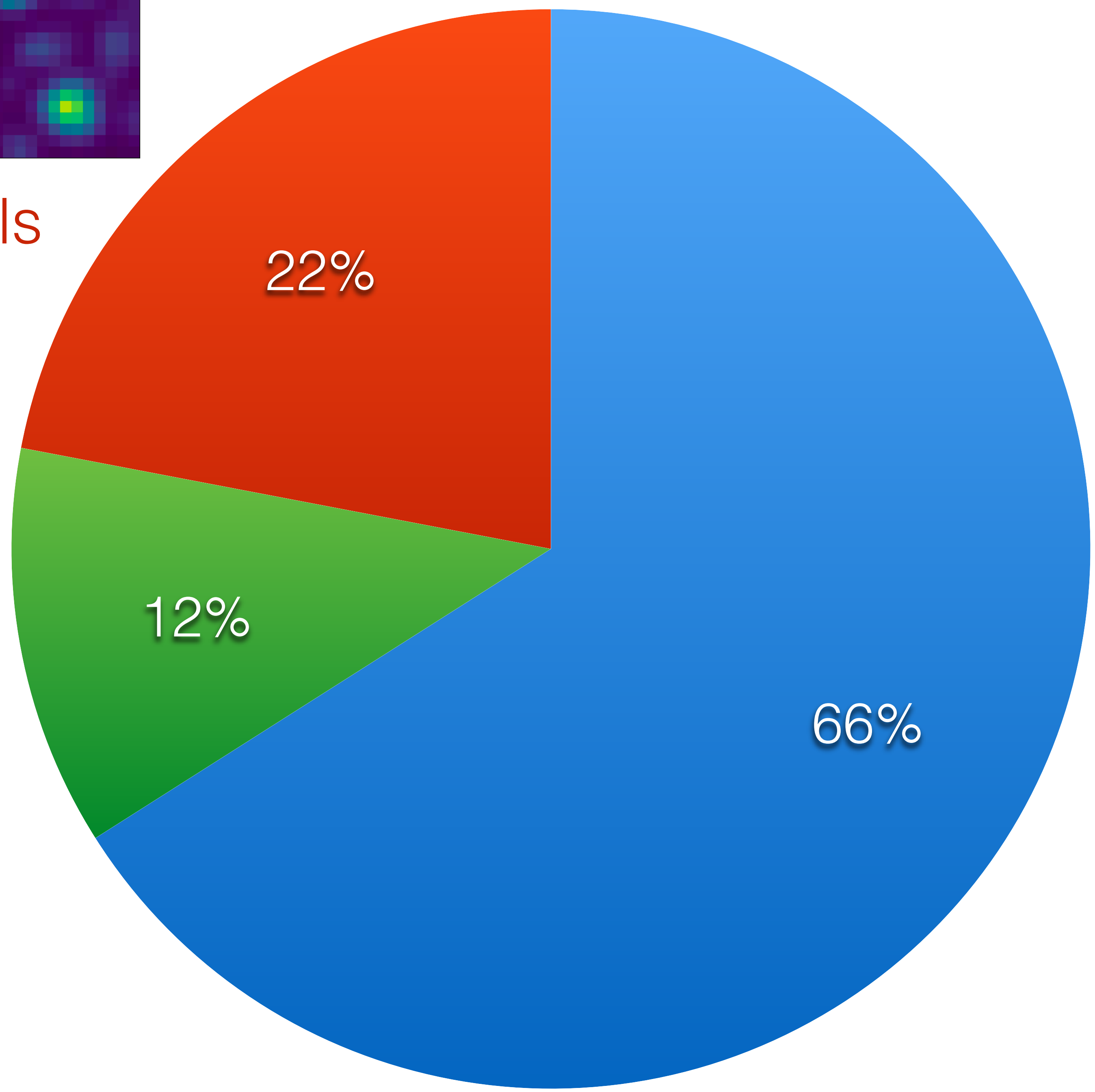
Moser et al. 2008

Accounting for heterogeneous code?

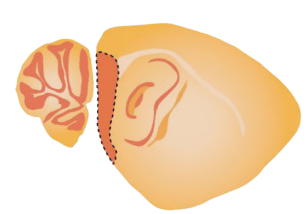
Accounting for heterogeneous code?



Grid Cells



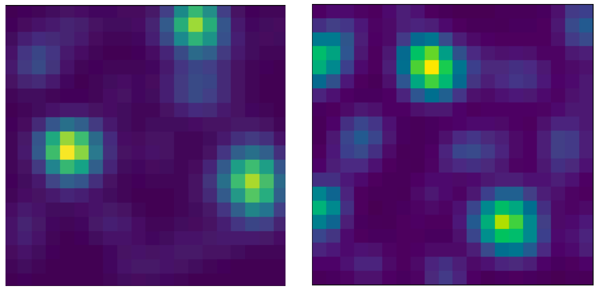
Data from: *Mallory et al. 2021*



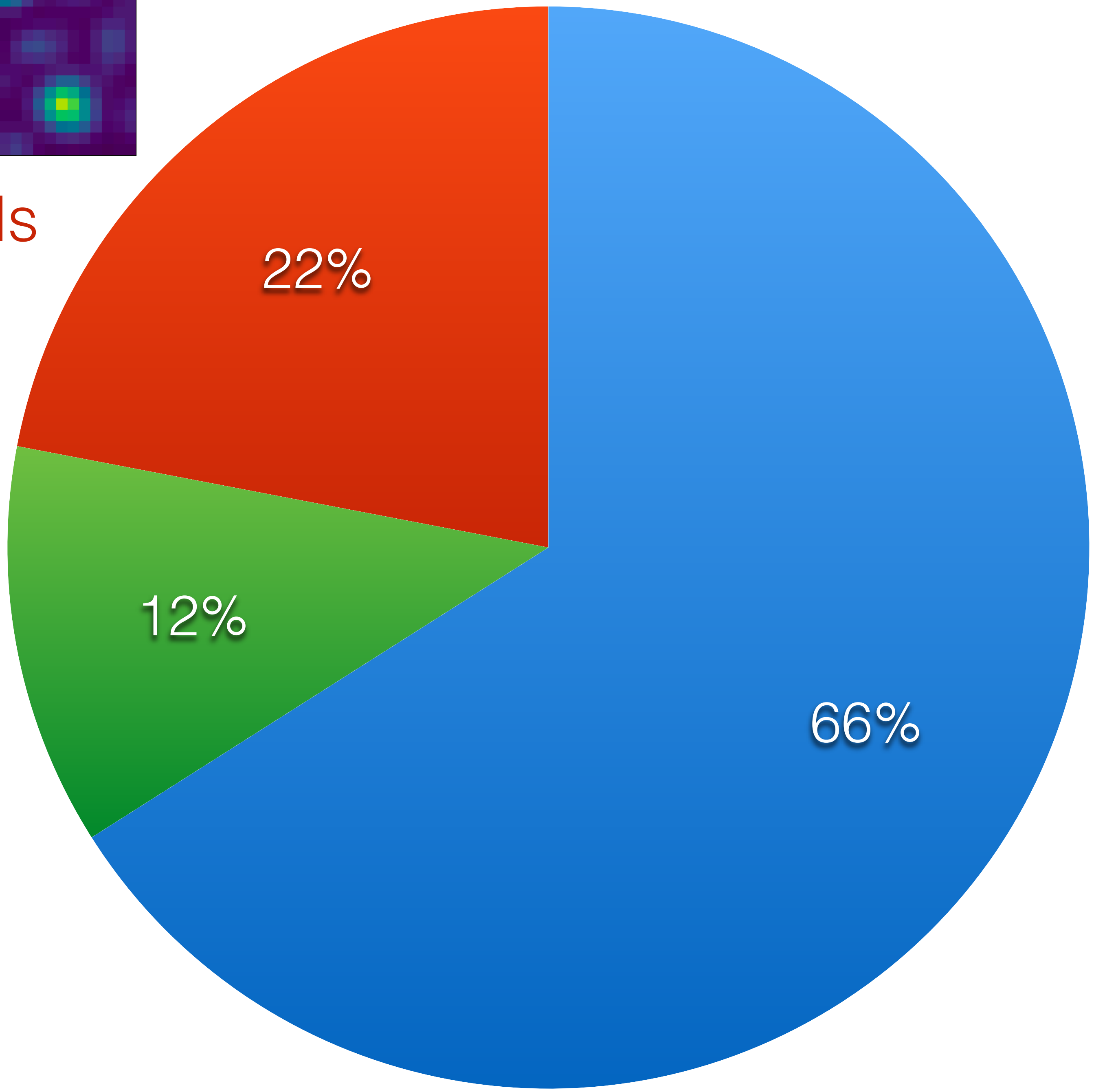
Caitlin Mallory

Accounting for heterogeneous code?

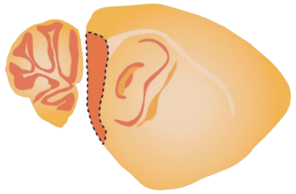
More like ~2-3%!



Grid Cells



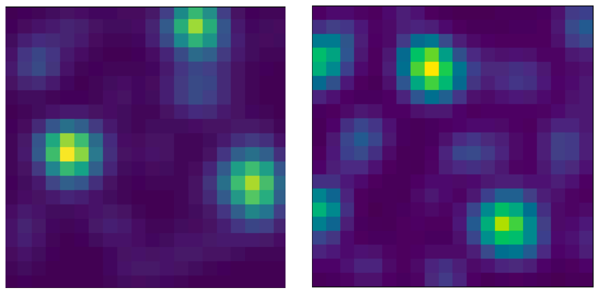
Data from: *Mallory et al. 2021*



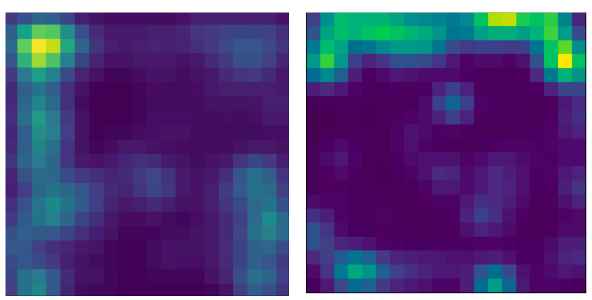
Caitlin Mallory

Accounting for heterogeneous code?

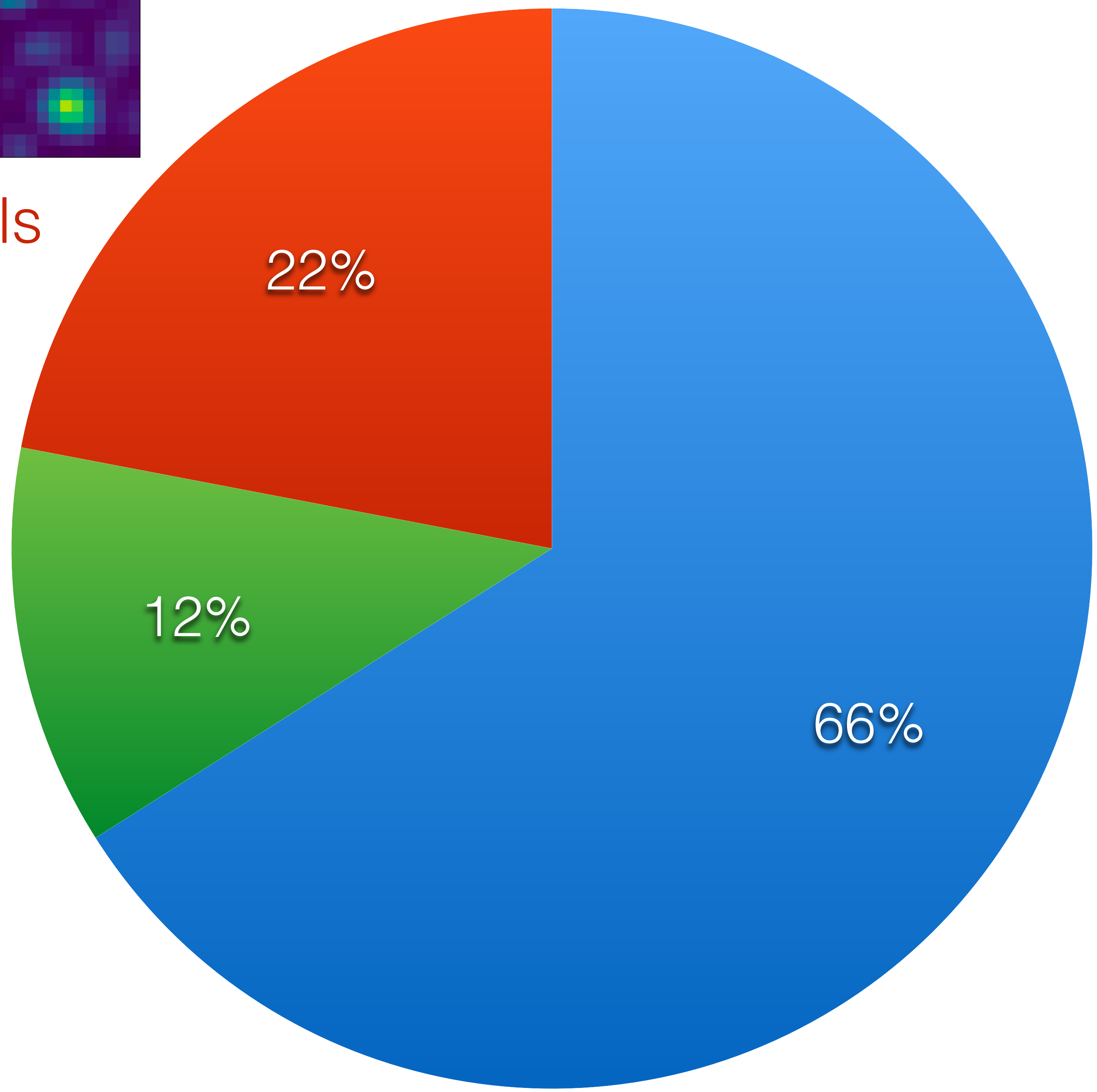
More like ~2-3%!



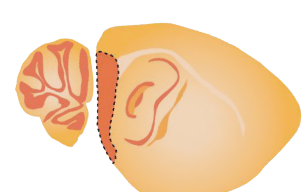
Grid Cells



Border Cells

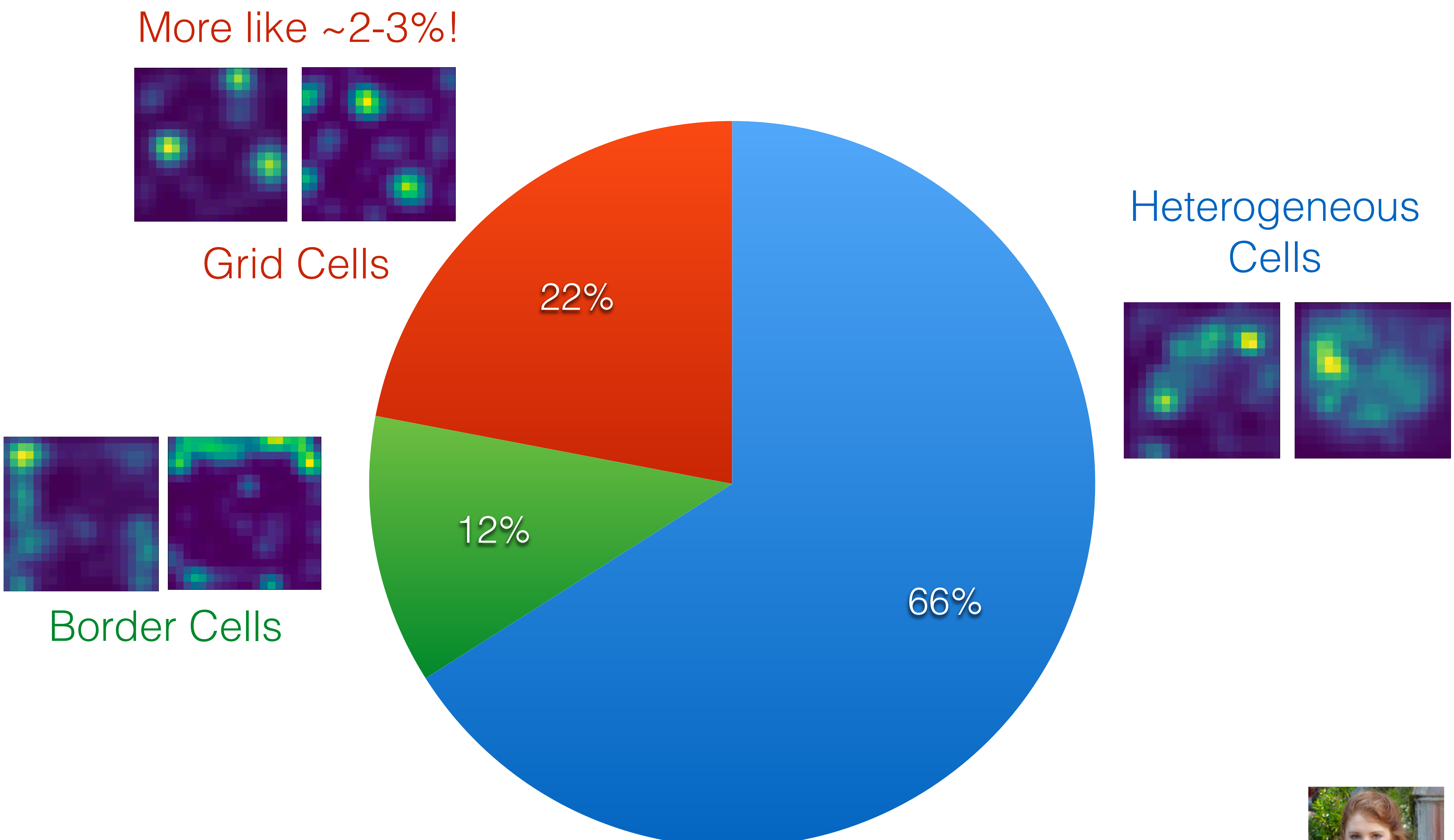


Data from: *Mallory et al. 2021*

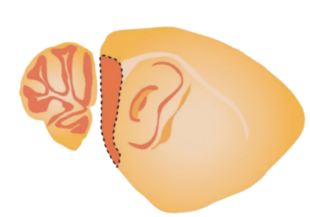


Caitlin Mallory

Accounting for heterogeneous code?

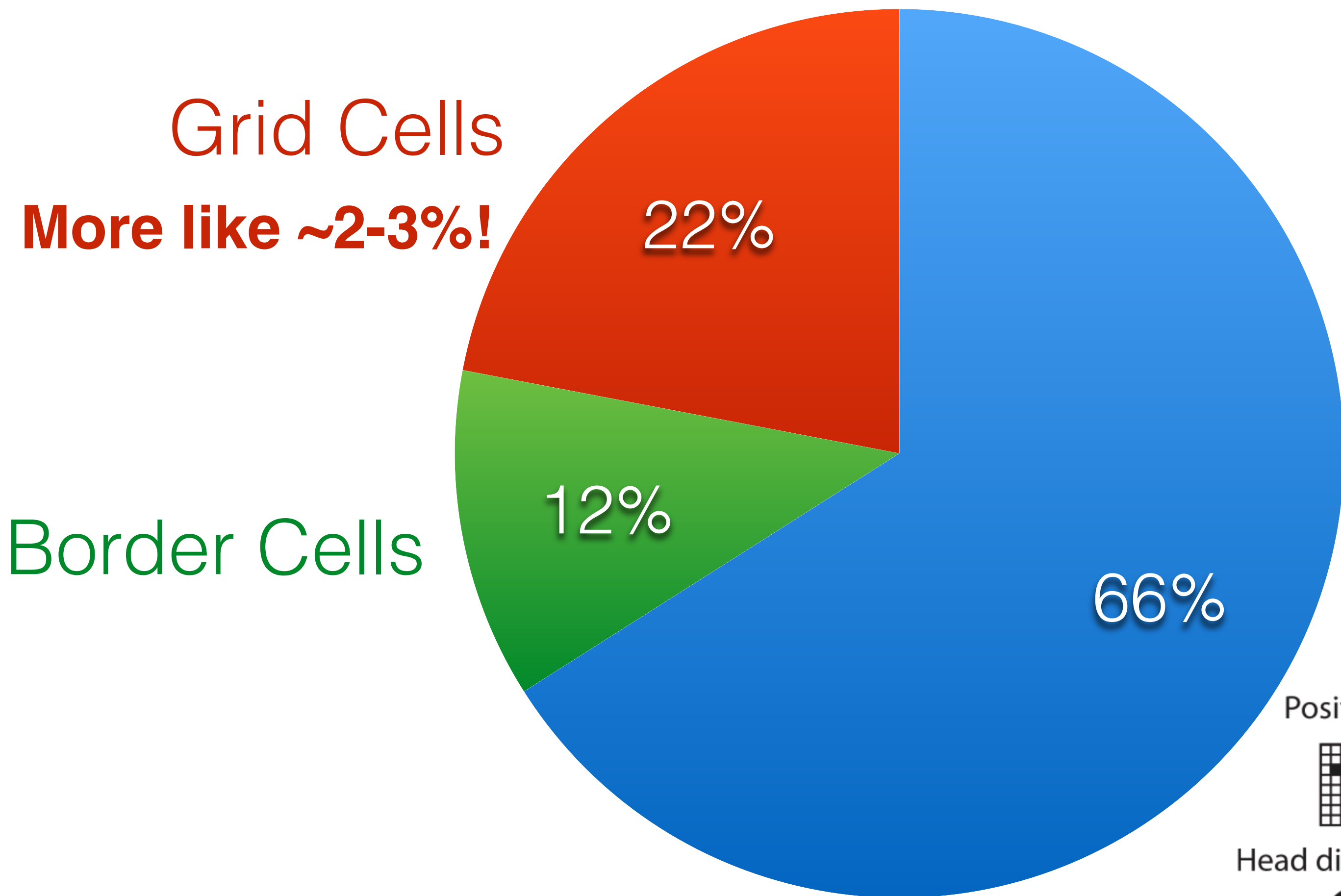


Data from: *Mallory et al. 2021*

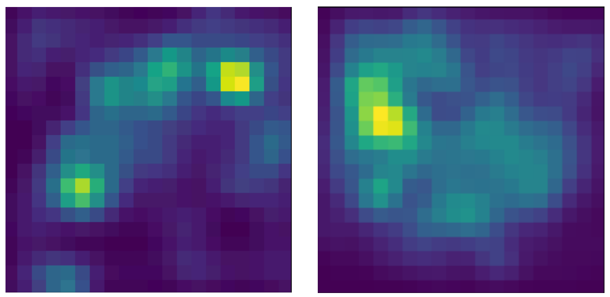


Caitlin Mallory

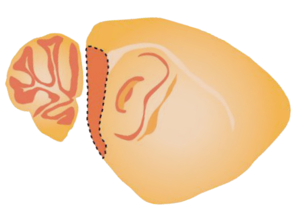
Accounting for heterogeneous code?

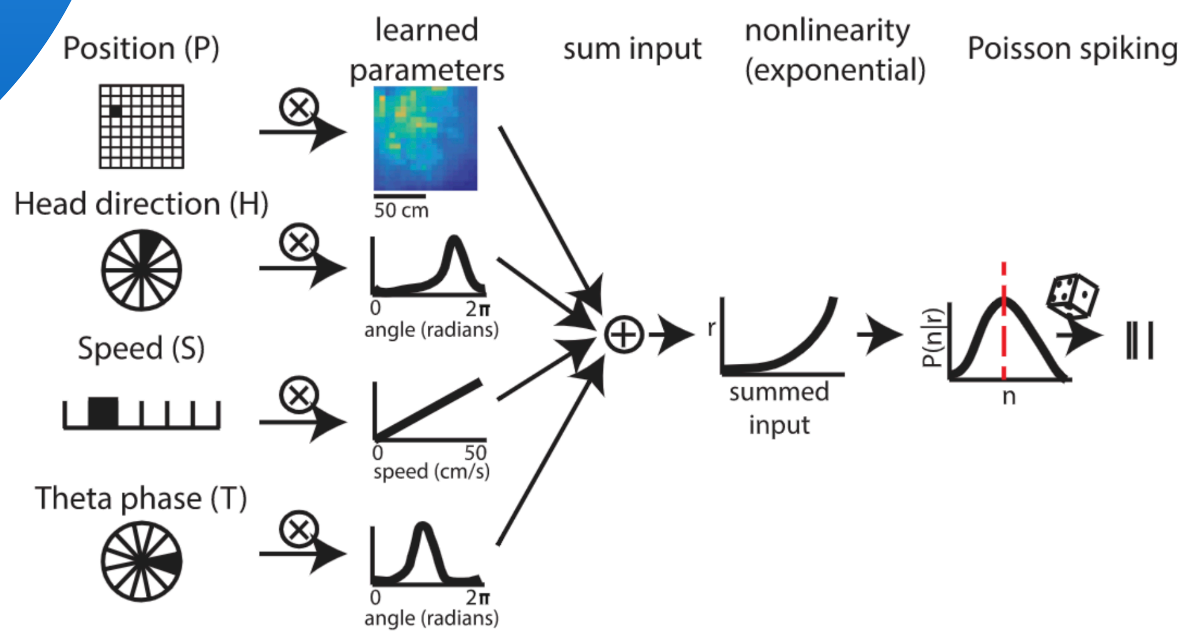


Heterogeneous
Cells



Data from: *Mallory et al. 2021*



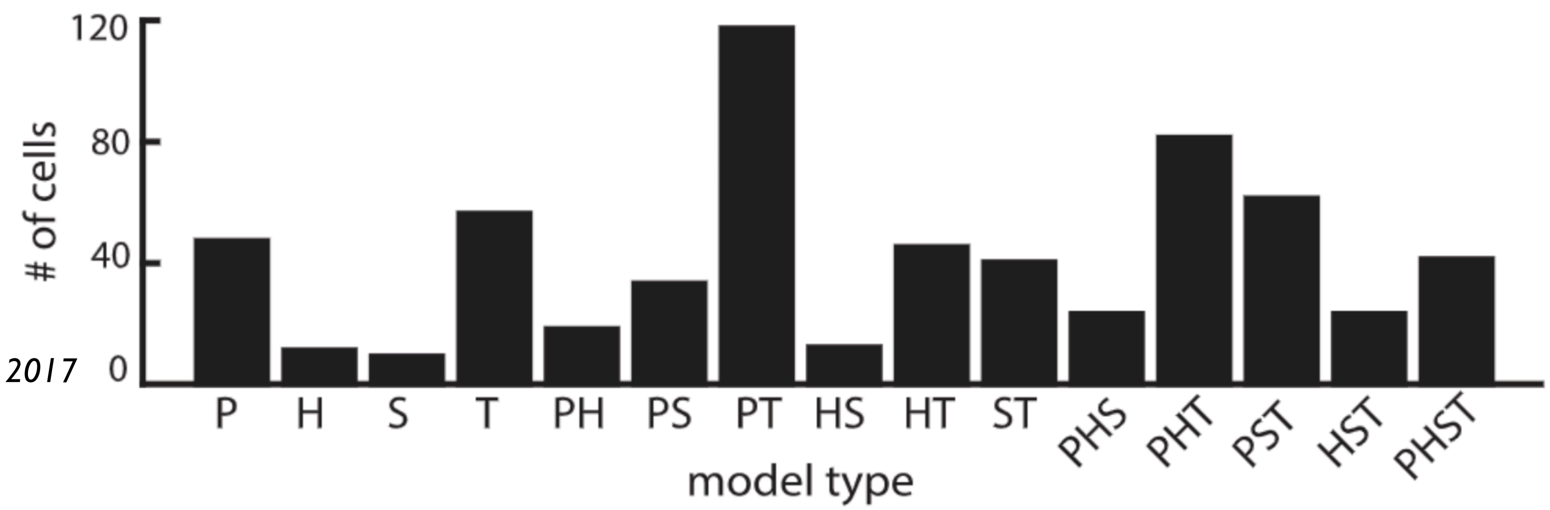



Kiah
Hardcastle

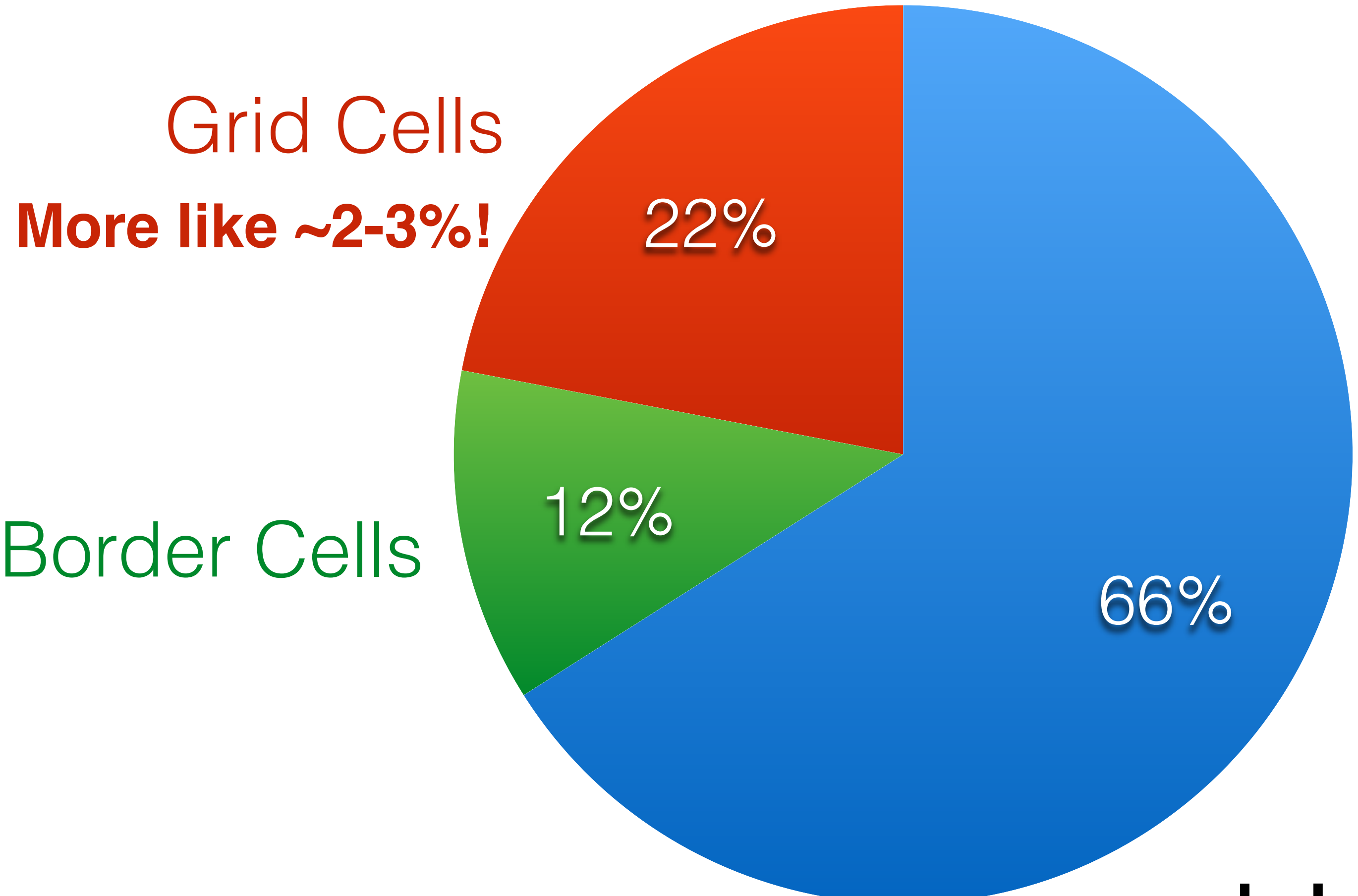

Surya Ganguli


Lisa Giocomo

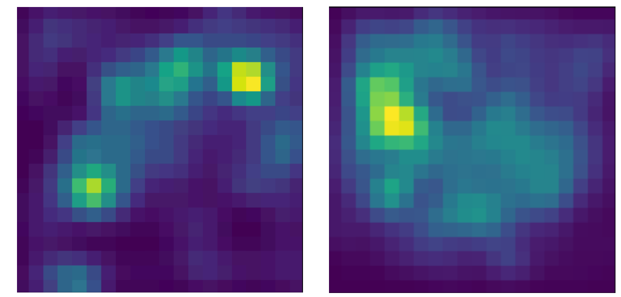
Hardcastle et al. 2017



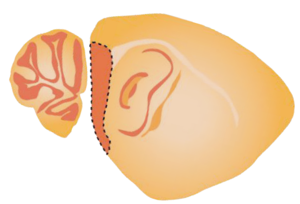
Accounting for heterogeneous code?



Heterogeneous Cells



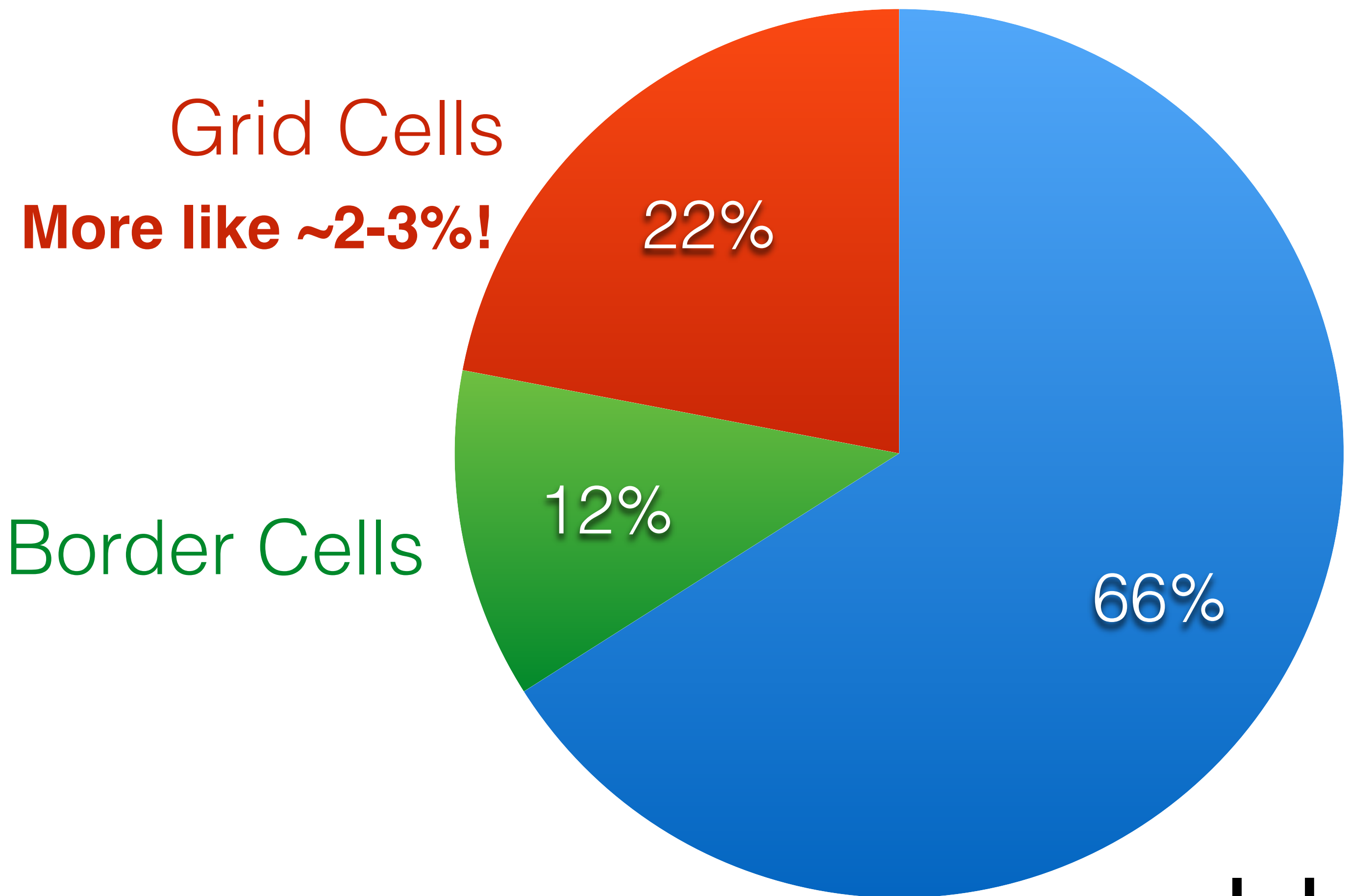
Data from: *Mallory et al. 2021*



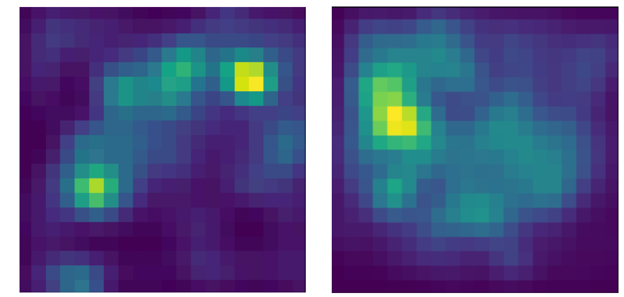
<u>Neurobiological Puzzle(s):</u>

1. How might we characterize what these heterogeneous cells do?

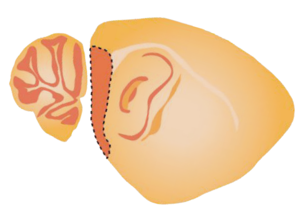
Accounting for heterogeneous code?



Heterogeneous Cells



Data from: *Mallory et al. 2021*

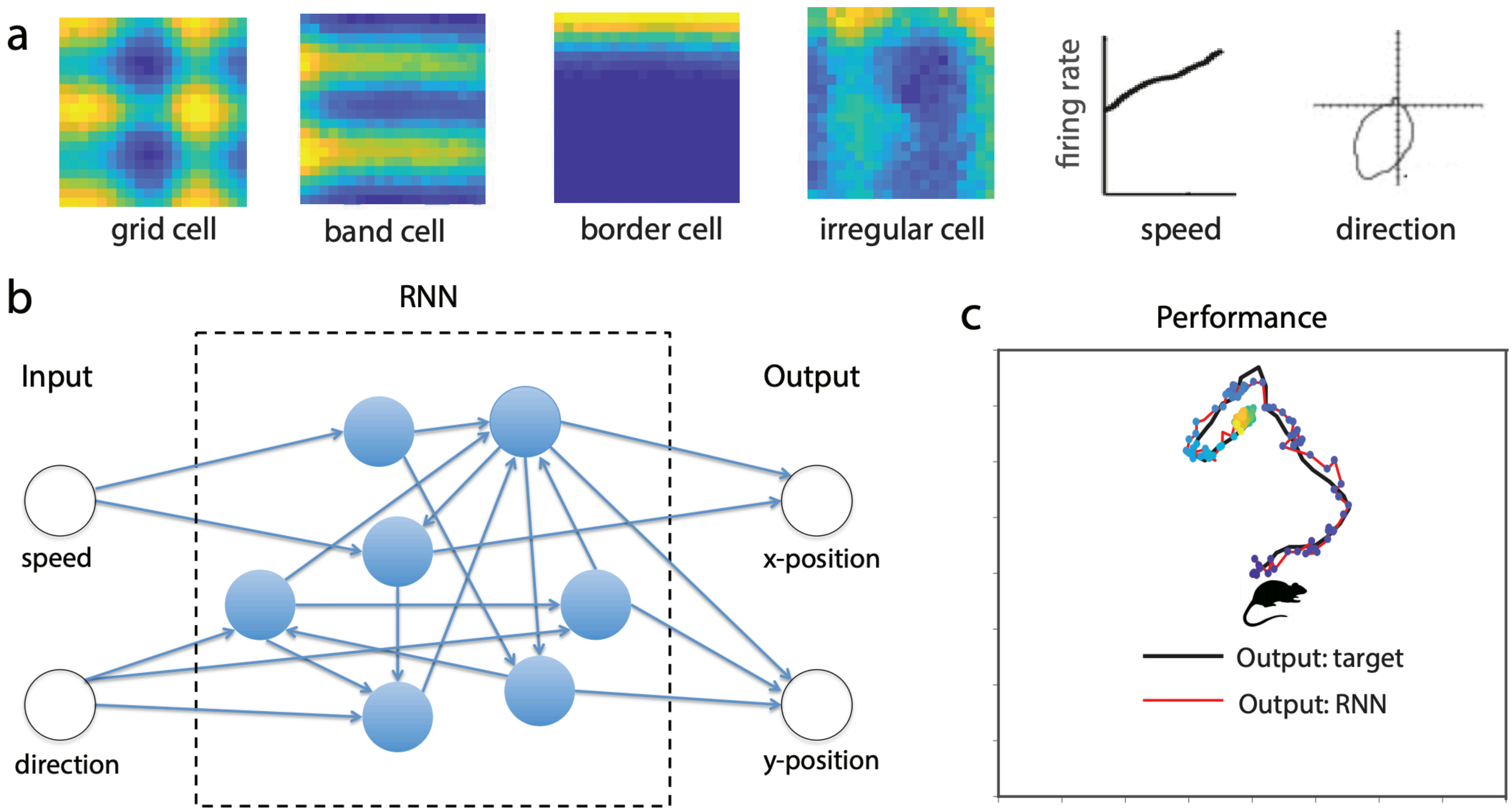


Neurobiological Puzzle(s):

1. How might we characterize what these heterogeneous cells do?

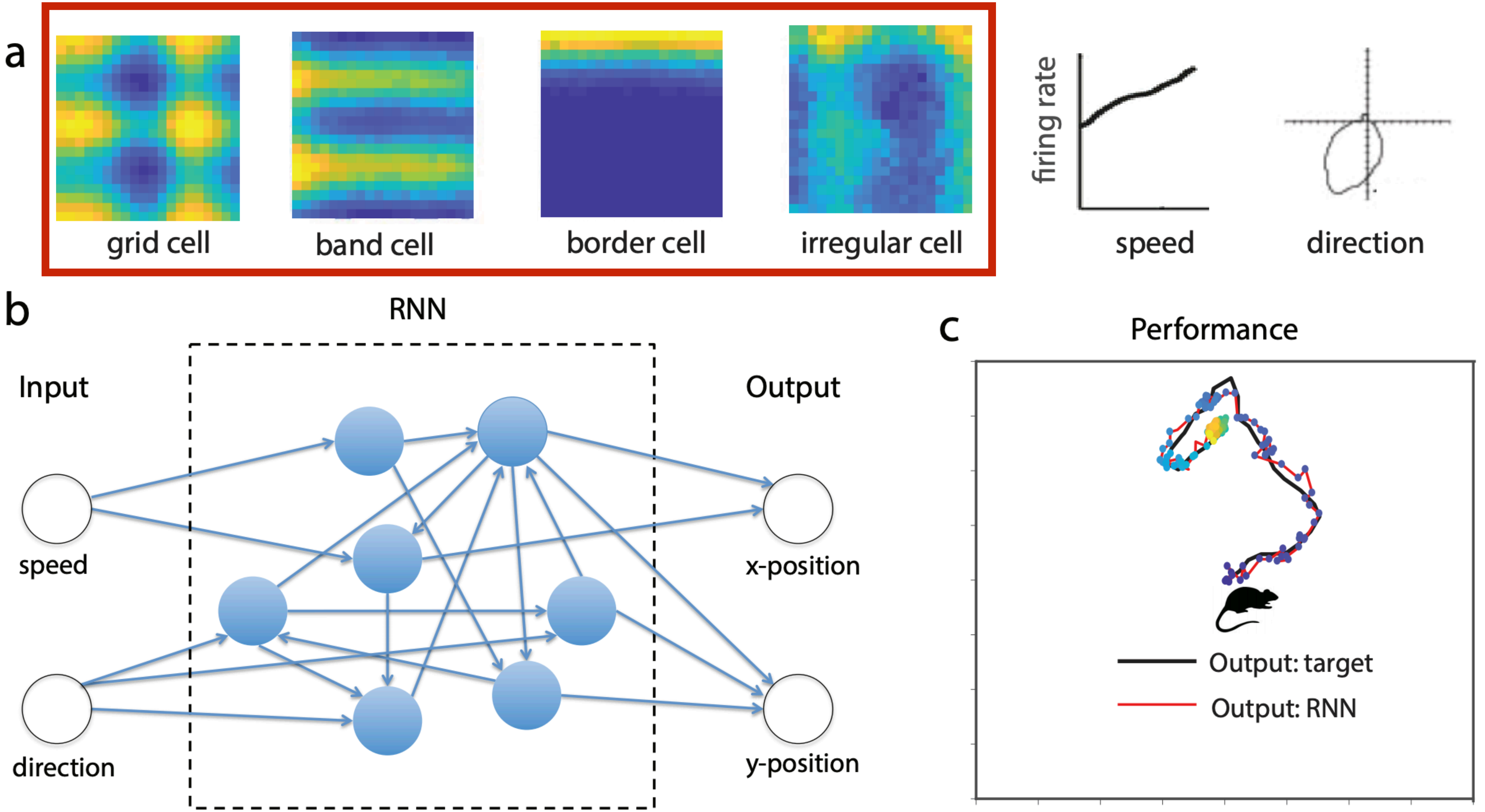
2. What functional role do these cells serve in the circuit, if any?

But more recently there are neural network models that "develop" these cells…



Cueva & Wei* 2018*

But more recently there are neural network models that "develop" these cells…

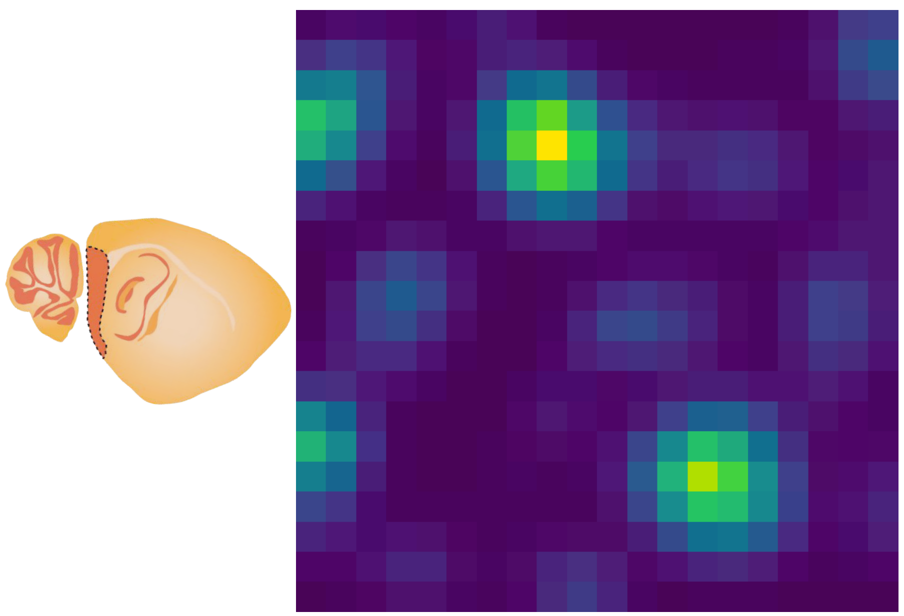


Cueva & Wei* 2018*

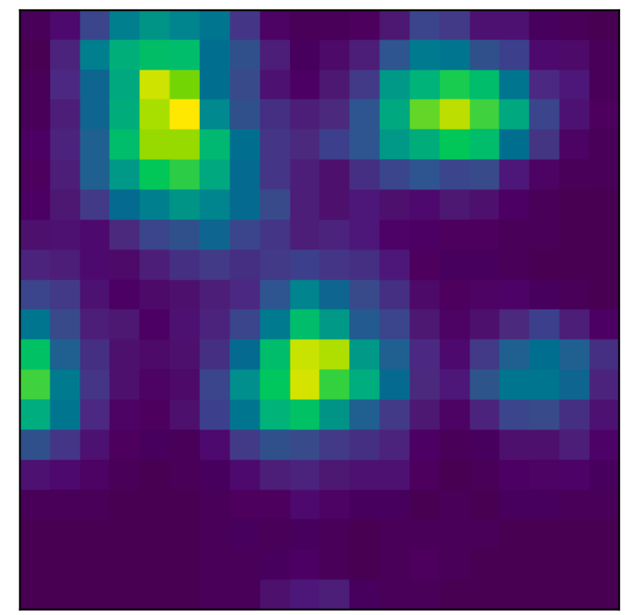
Goal-Driven Approach

But are they a good ***quantitative*** model of these responses?

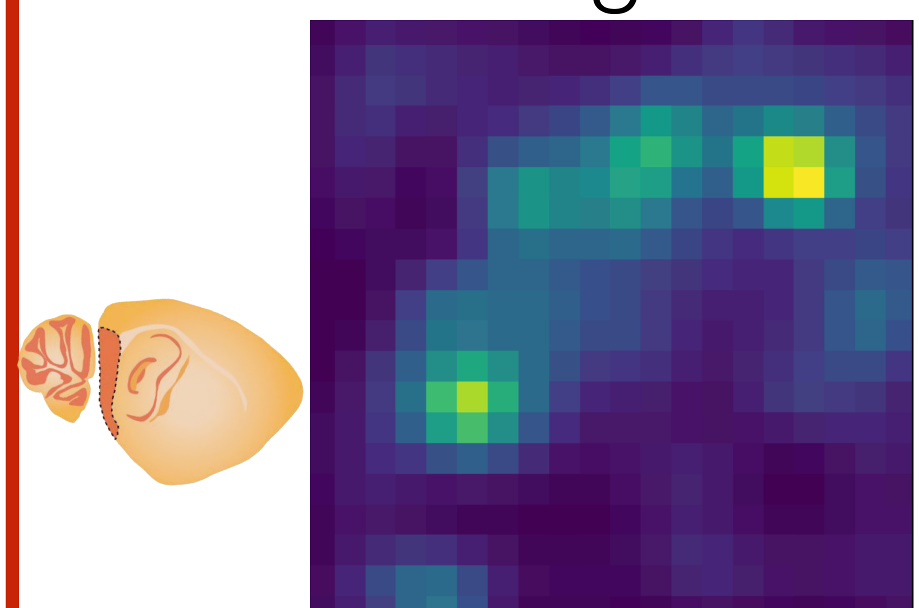
MEC Grid Cell



Model Grid Cell

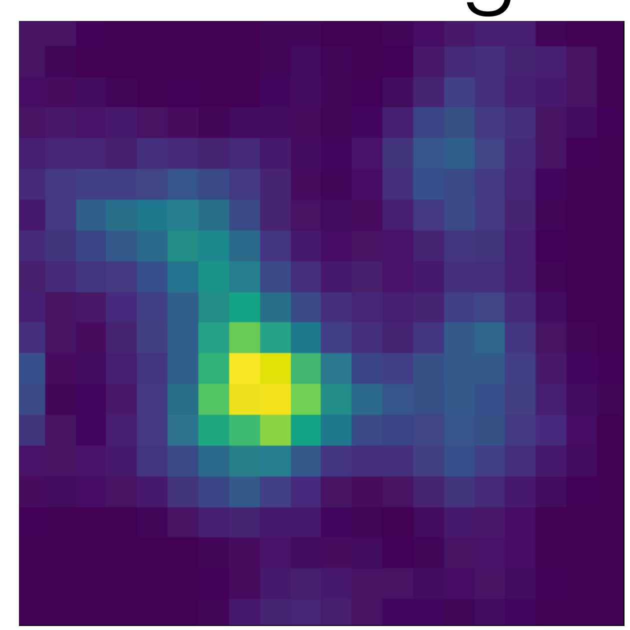


MEC Heterogeneous Cell



?

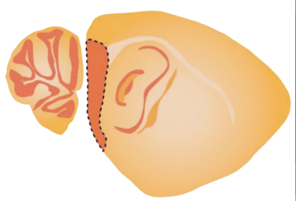
Model Heterogeneous Cell

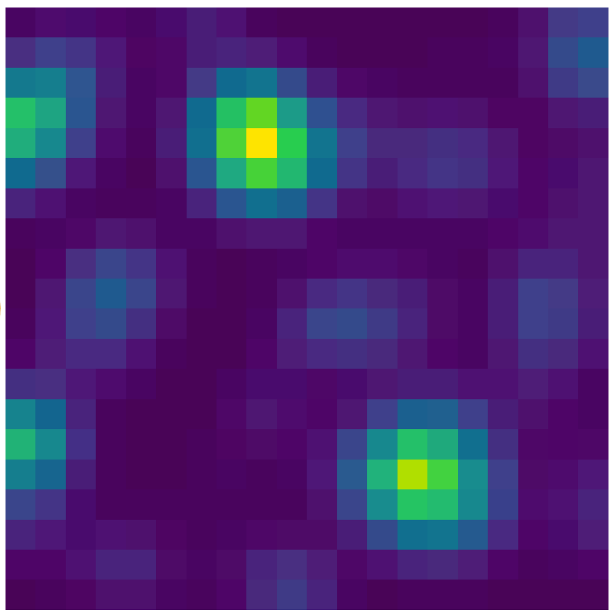


Goal-Driven Approach

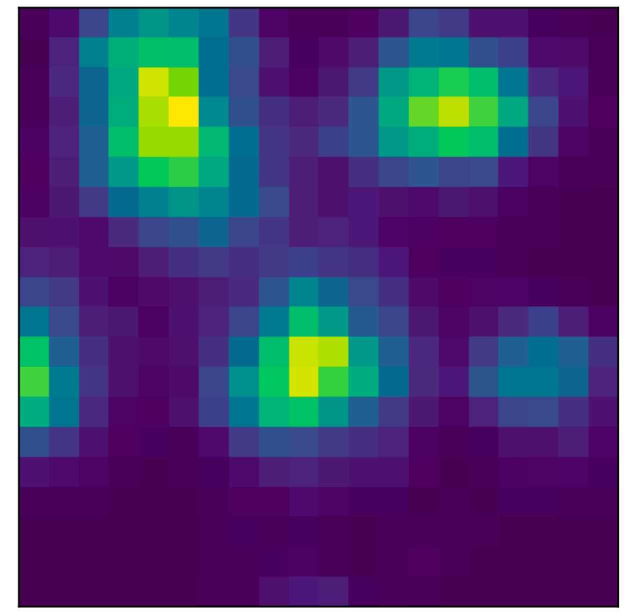
But are they a good ***quantitative*** model of these responses?

MEC Grid Cell

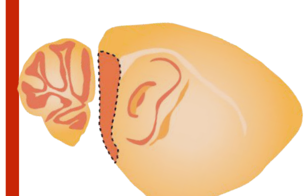


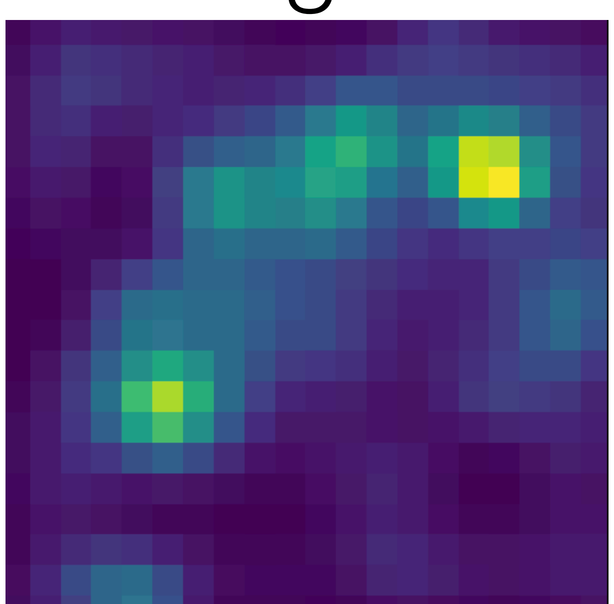


Model Grid Cell



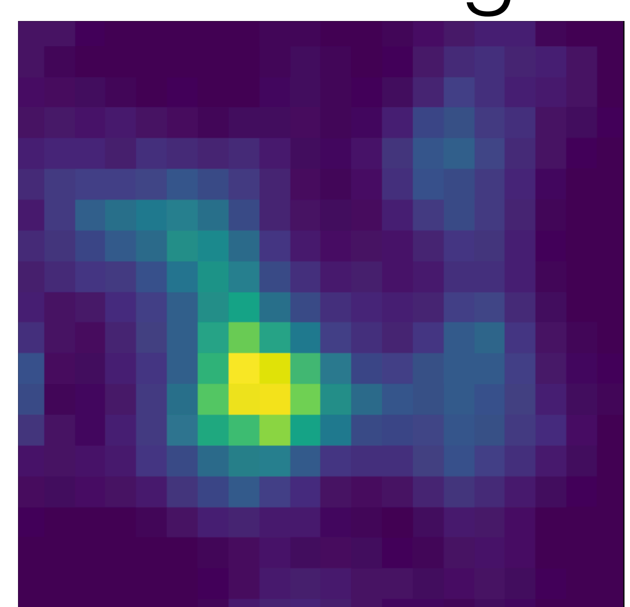
MEC Heterogeneous Cell





Not all models are equal!

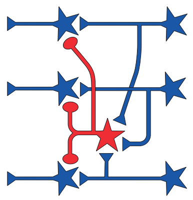
Model Heterogeneous Cell



Goal-Driven Approach

A = *architecture class*

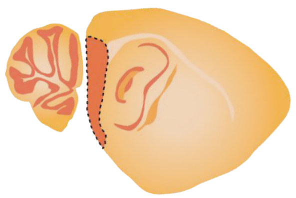
1. *"Circuit"*

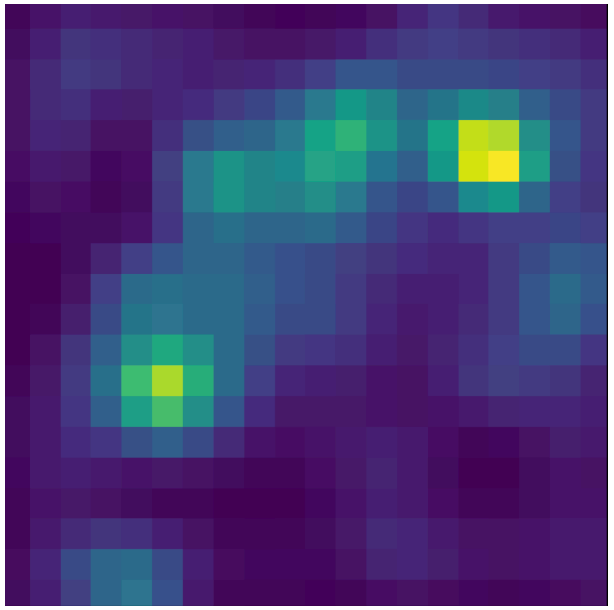


T = *task loss*

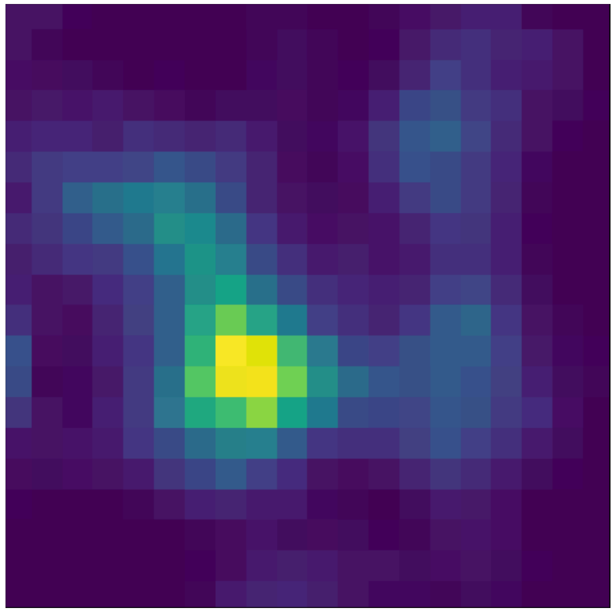
3. *"Ecological niche/behavior"*

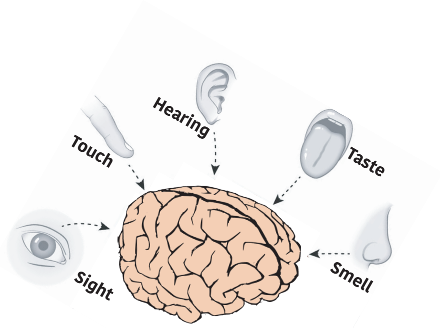
MEC Heterogeneous Cell





Model Heterogeneous Cell





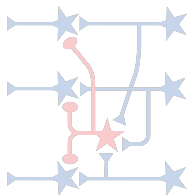
2. *"Environment"*

D = *data stream*

Goal-Driven Approach

A = *architecture class*

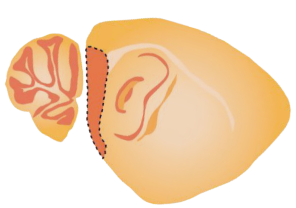
1. *"Circuit"*

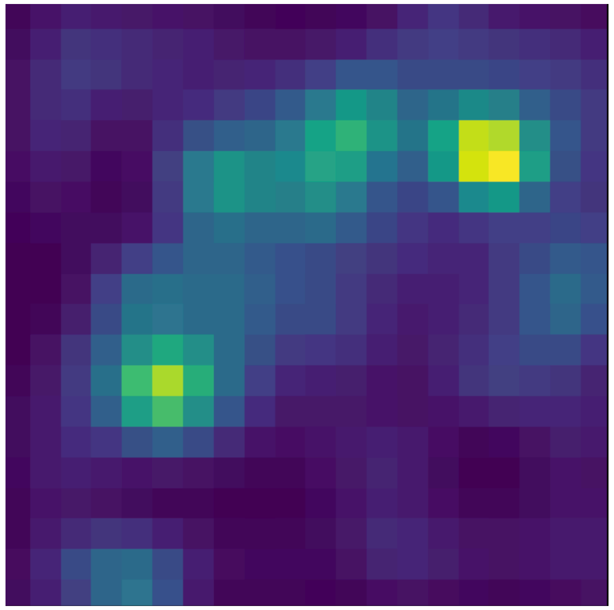


T = *task loss*

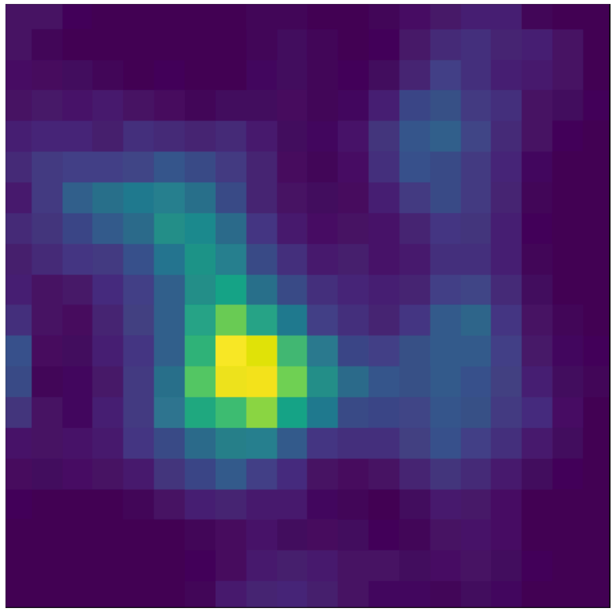
3. *"Ecological niche/behavior"*

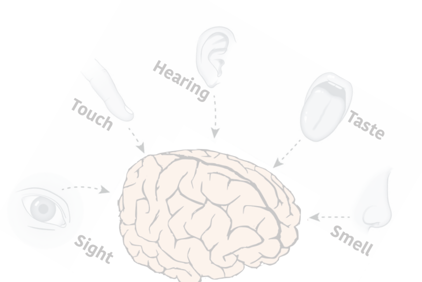
MEC Heterogeneous Cell





Model Heterogeneous Cell

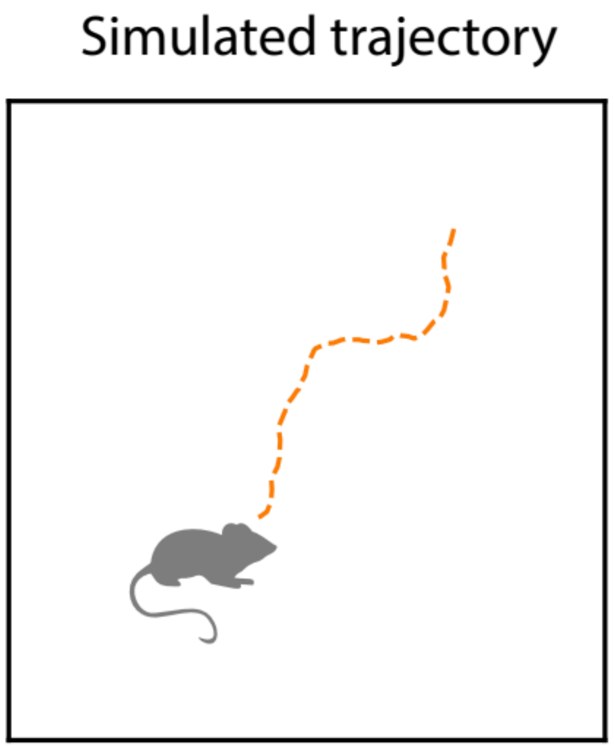


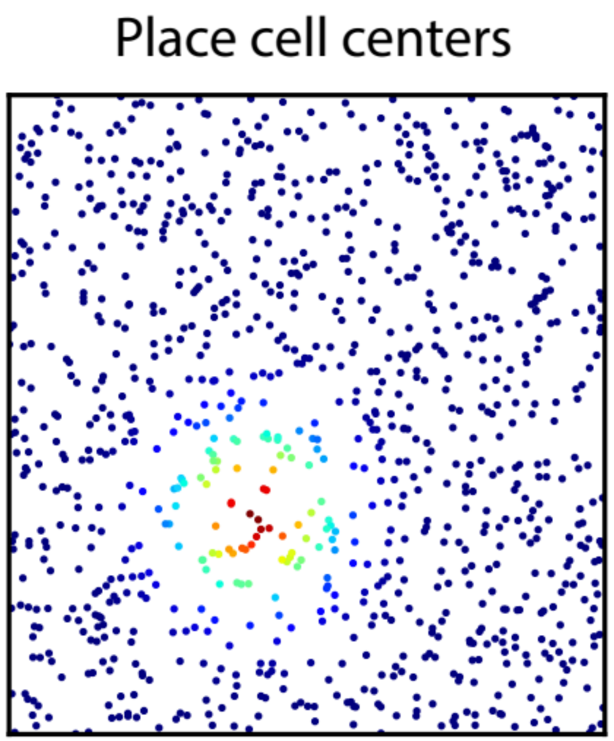


2. *"Environment"*

D = *data stream*

A spectrum of tasks

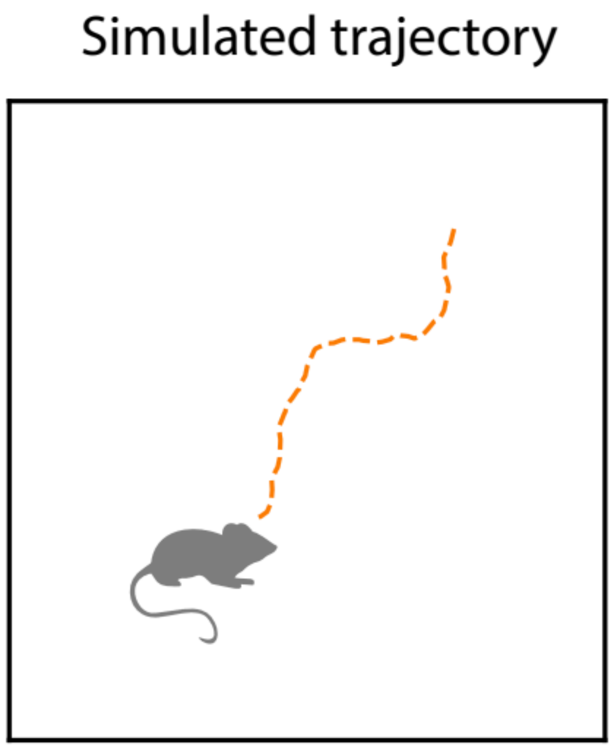


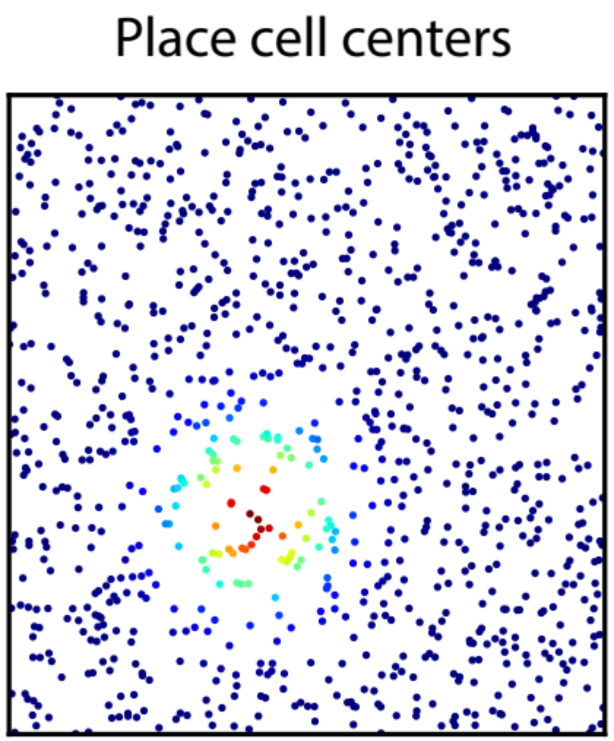


Banino, Barry* et al. 2018*

Sorscher, Mel* et al. 2019*

A spectrum of tasks





Banino, Barry* et al. 2018*

Sorscher, Mel* et al. 2019*

Velocity → MEC → Place Cells → Position (x,y)

A spectrum of tasks

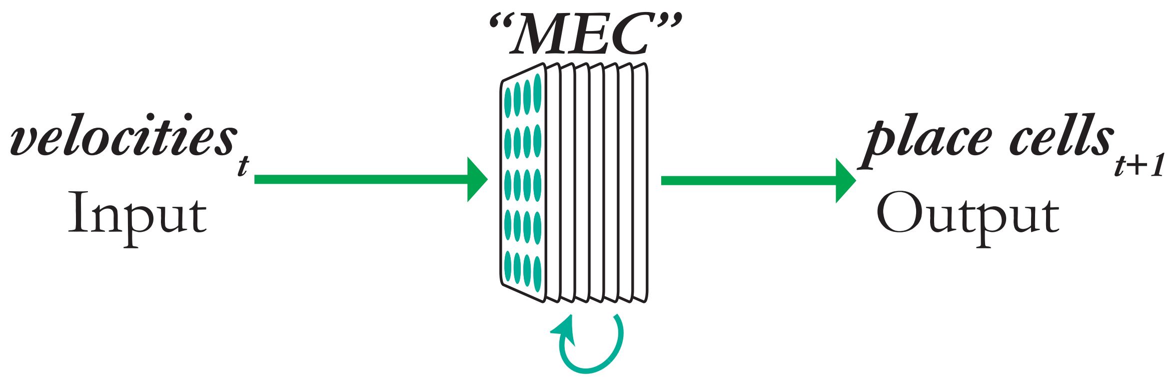
Simplest “model”

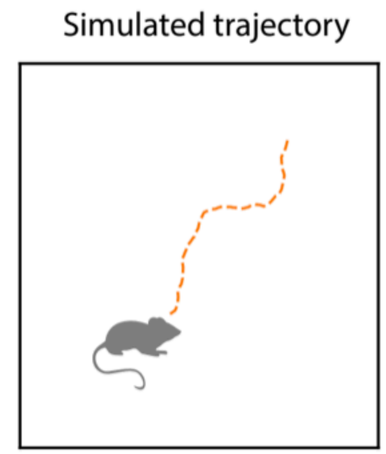
$\boxed{\text{Velocity}} \longrightarrow \text{MEC} \longrightarrow \text{Place Cells} \longrightarrow \text{Position } (x,y)$

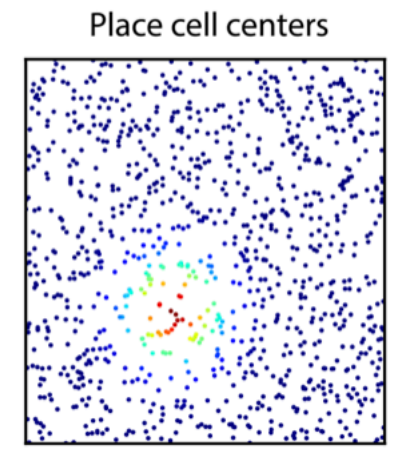
A spectrum of tasks

$$\mathcal{L}(\hat{p}, p) := -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N_p} p_i^t \log \hat{p}_i^t$$

Banino, Barry* et al. 2018*





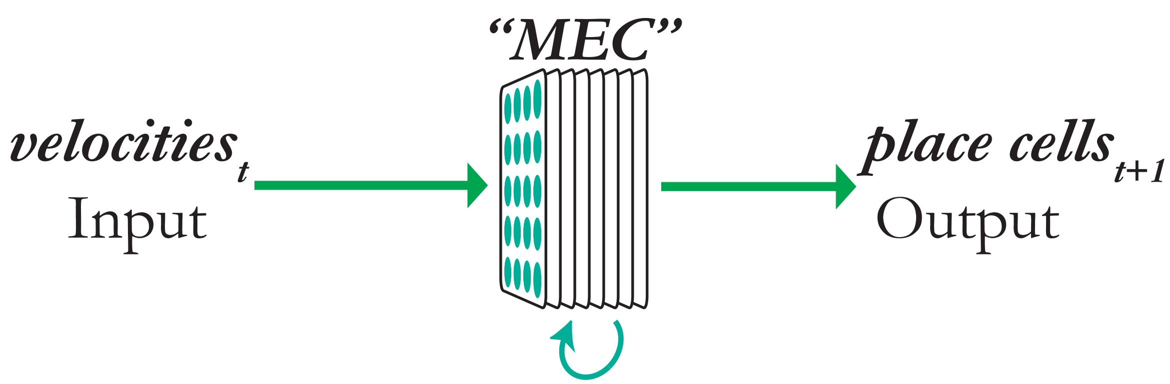


Velocity ⟶ MEC ⟶ Place Cells ⟶ Position (x,y)

A spectrum of tasks

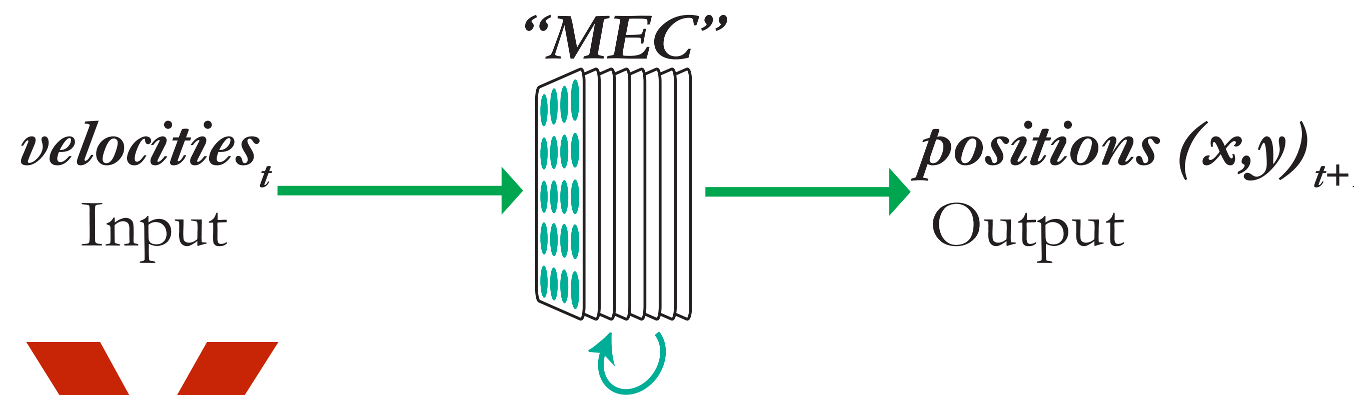
Banino, Barry* et al. 2018*

$$\mathcal{L}(\hat{p}, p) := -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N_p} p_i^t \log \hat{p}_i^t$$



Cueva & Wei* 2018*

$$\mathcal{L}(\hat{p}, p) := \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left(\left(p_x^t - \widehat{p}_x^t\right)^2 + \left(p_y^t - \widehat{p}_y^t\right)^2\right)$$

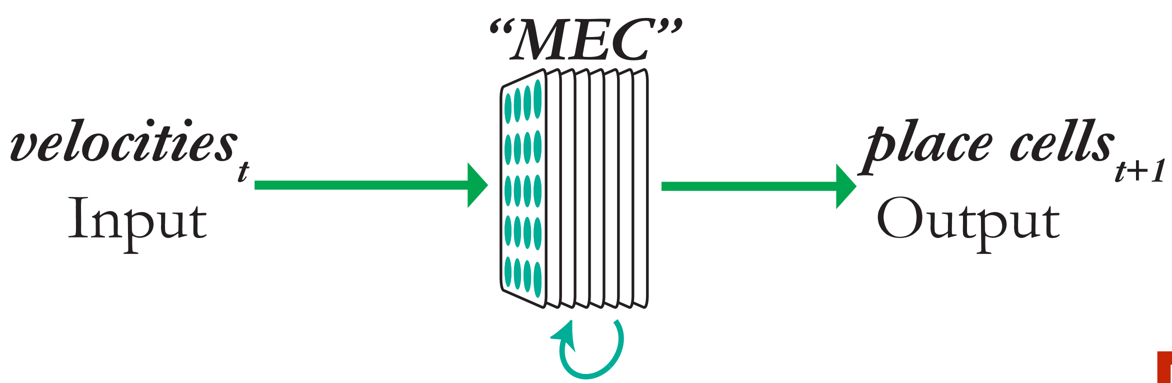


Velocity $\longrightarrow$ MEC $\longrightarrow$ Place Cells $\longrightarrow$ Position (x,y)

A spectrum of tasks

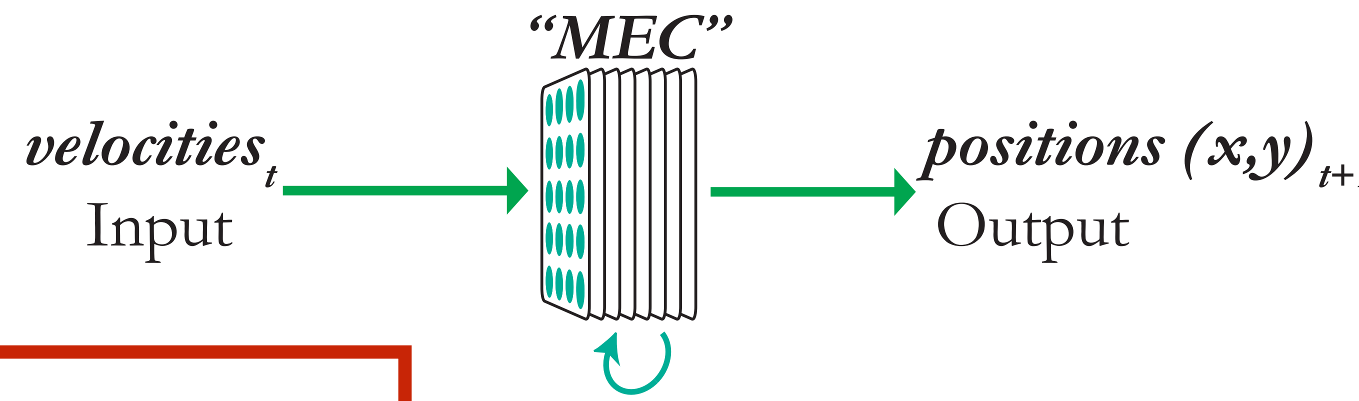
Banino, Barry* et al. 2018*

$$\mathcal{L}(\hat{p}, p) := -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N_p} p_i^t \log \hat{p}_i^t$$



Cueva & Wei* 2018*

$$\mathcal{L}(\hat{p}, p) := \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left(\left(p_x^t - \hat{p}_x^t\right)^2 + \left(p_y^t - \hat{p}_y^t\right)^2\right)$$



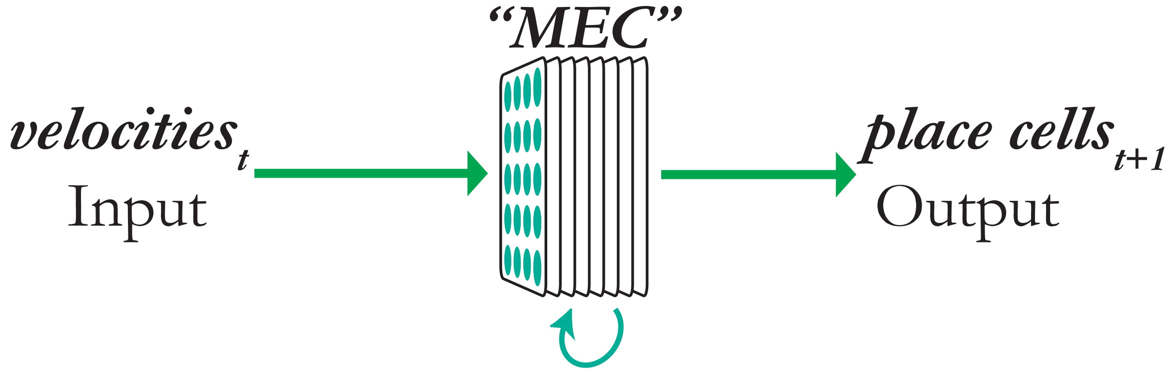
Velocity → MEC → Place Cells → Position (x,y)

Output-based models

A spectrum of tasks

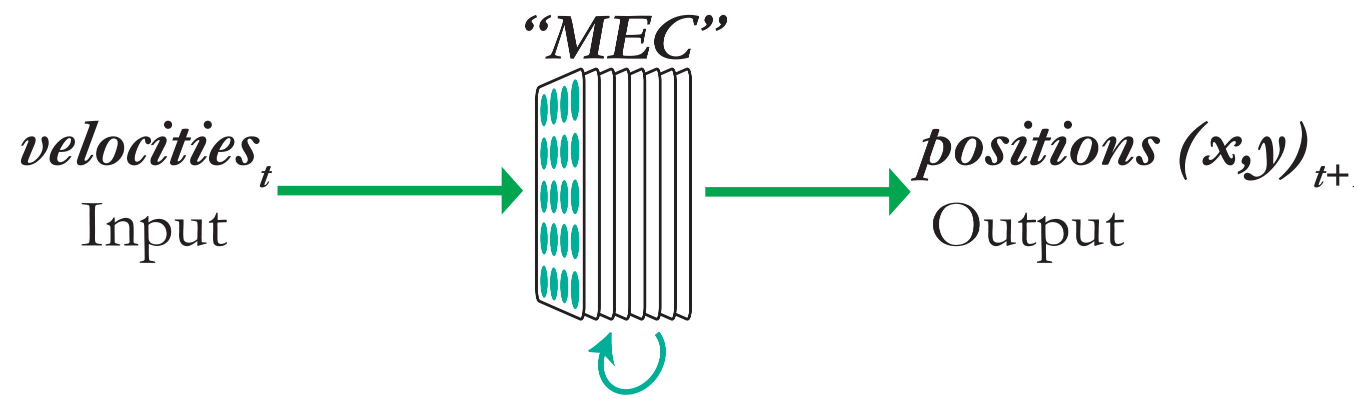
Banino, Barry* et al. 2018*

$$\mathcal{L}(\hat{p}, p) := -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N_p} p_i^t \log \hat{p}_i^t$$



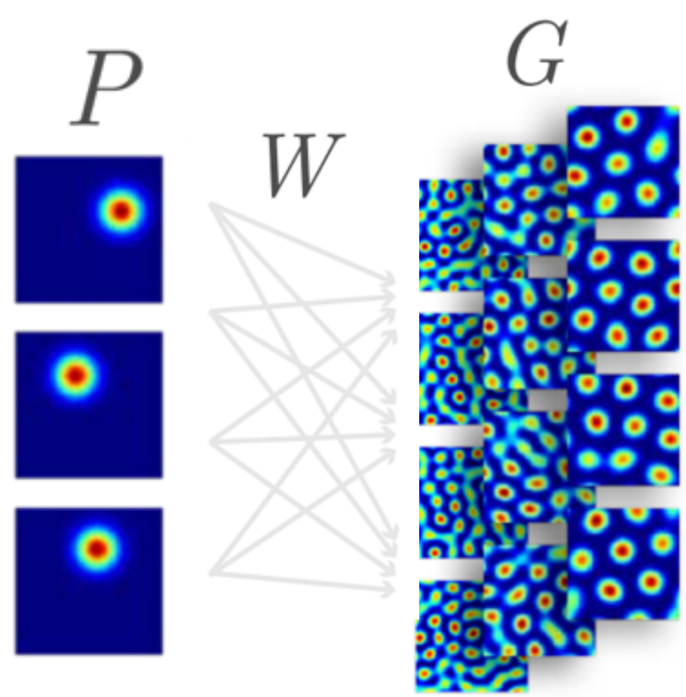
Cueva & Wei* 2018*

$$\mathcal{L}(\hat{p}, p) := \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left(\left(p_x^t - \hat{p}_x^t\right)^2 + \left(p_y^t - \hat{p}_y^t\right)^2\right)$$



Velocity $\longrightarrow$ MEC $\longrightarrow$ Place Cells $\longrightarrow$ Position (x,y)

NMF
(Place Cell Input)

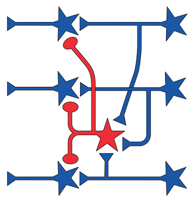


Dordek et al. 2016

Goal-Driven Approach

A = *architecture class*

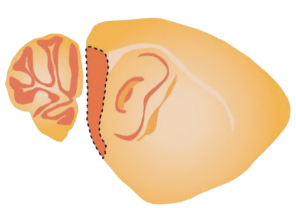
1. *"Circuit"*

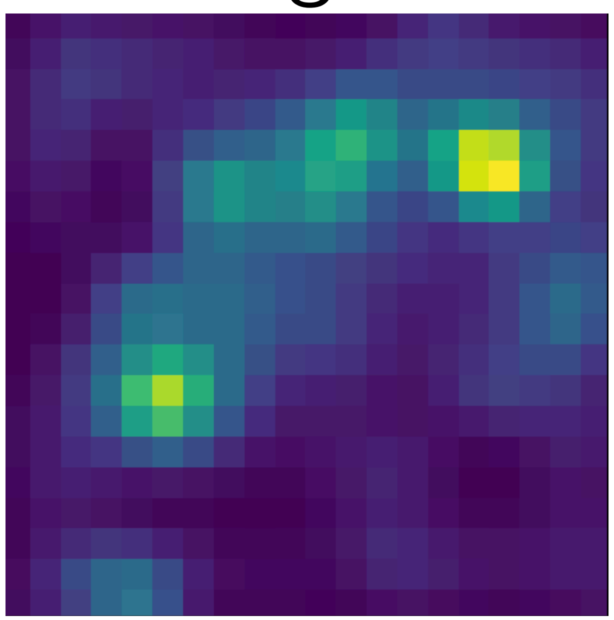


T = *task loss*

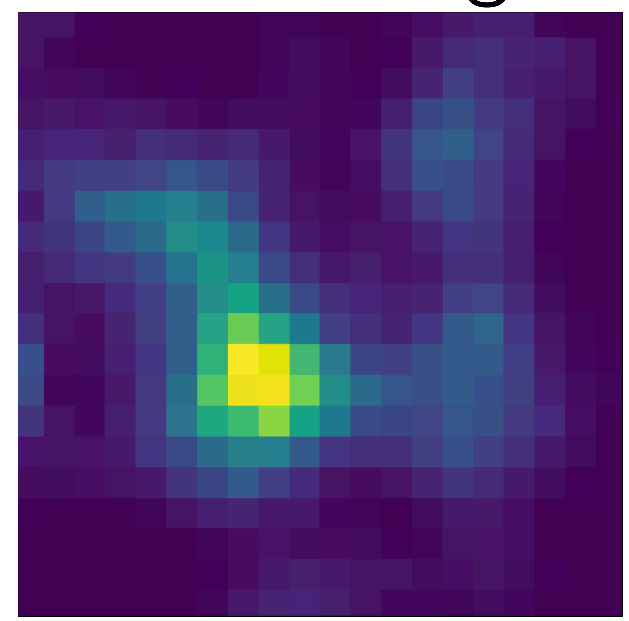
3. *"Ecological niche/behavior"*

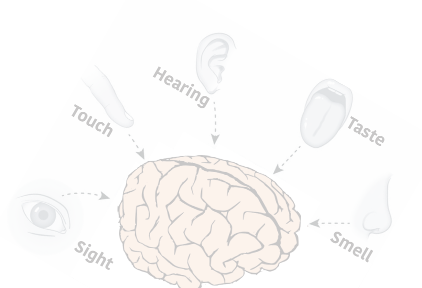
MEC Heterogeneous Cell





Model Heterogeneous Cell

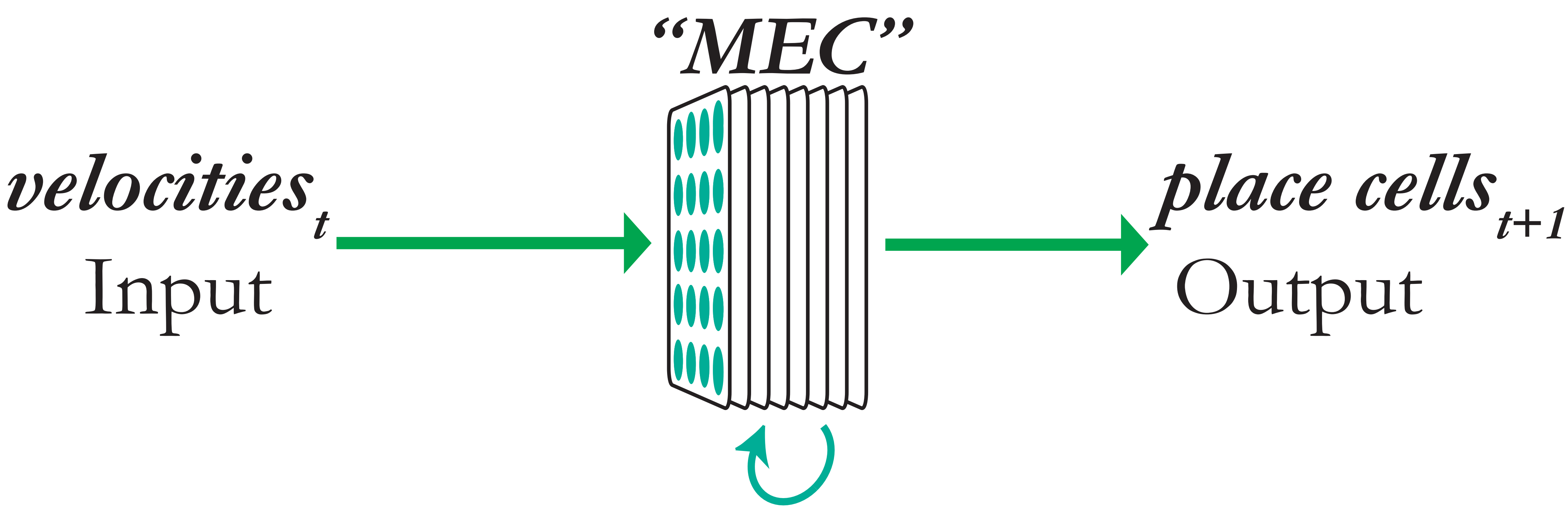




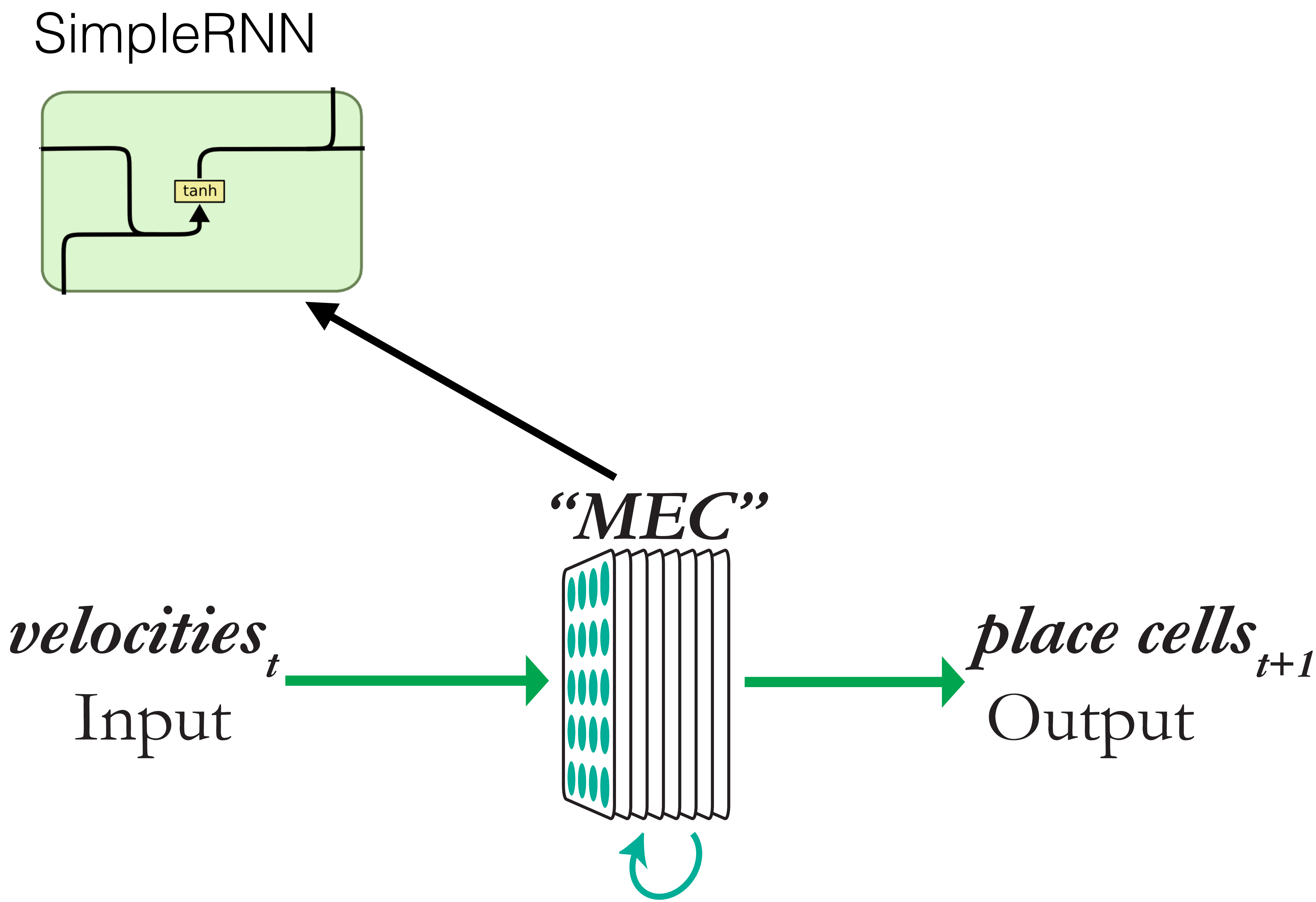
2. *"Environment"*

D = *data stream*

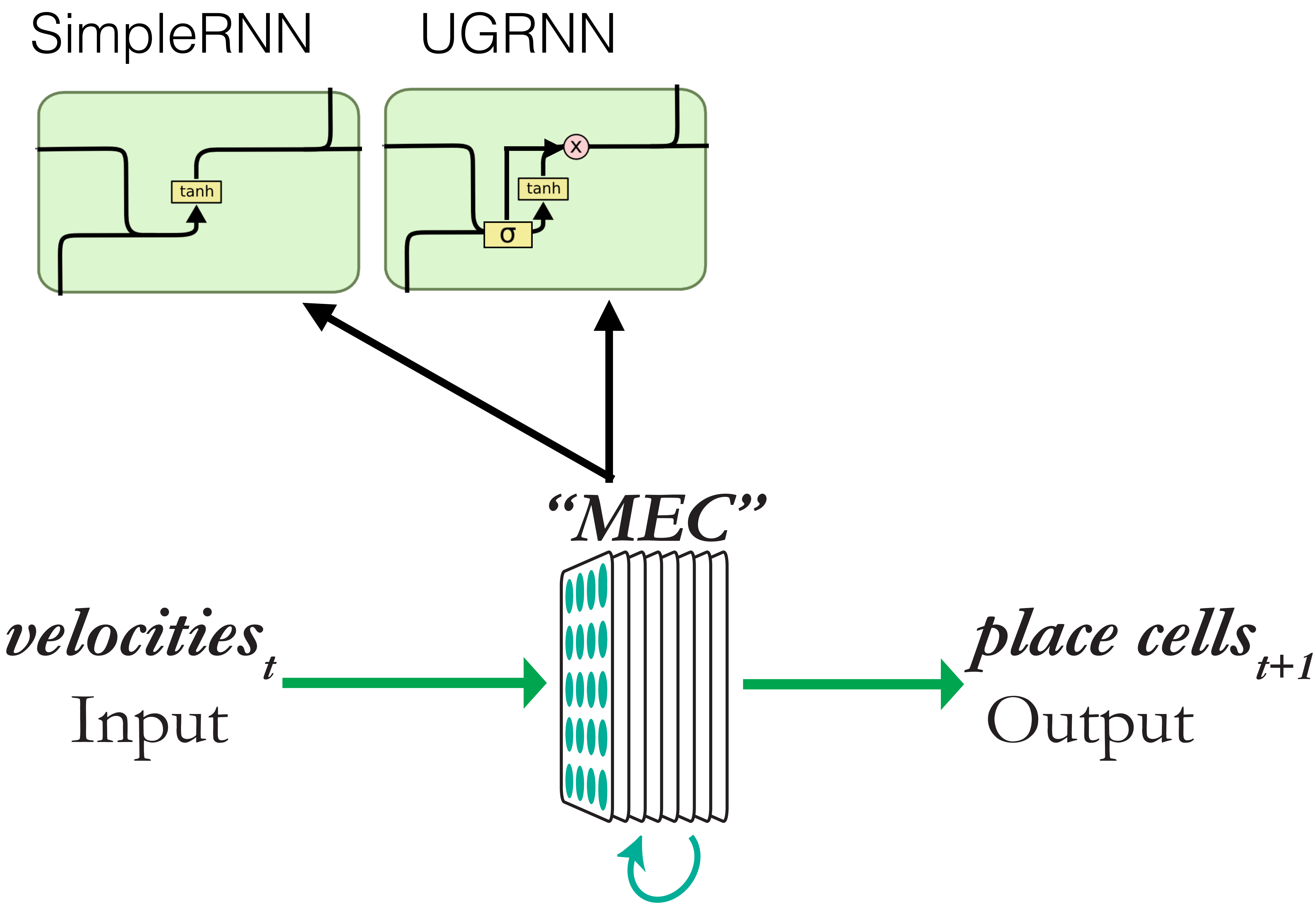
A spectrum of circuits



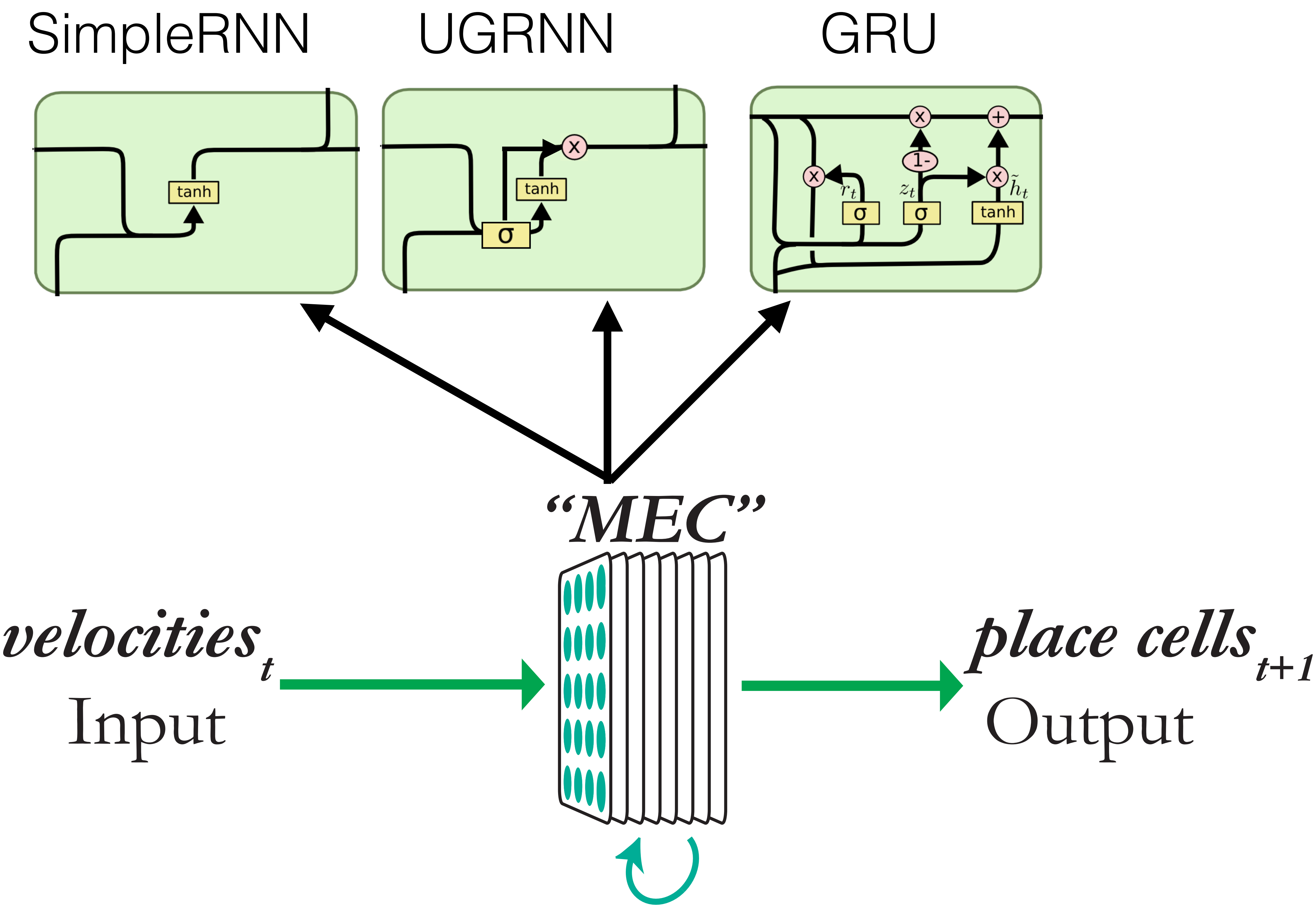
A spectrum of circuits



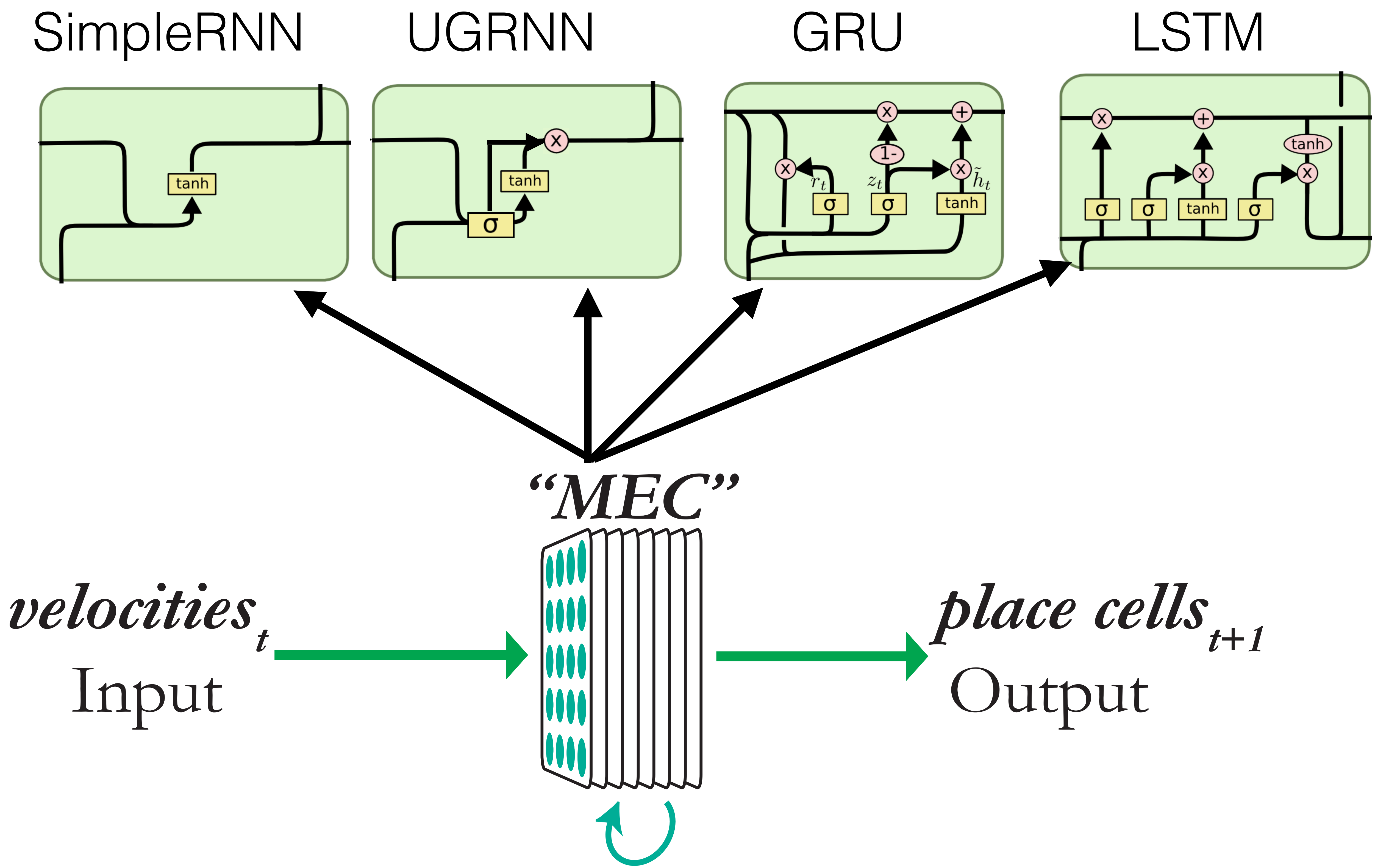
A spectrum of circuits — learnable modulation (“gating”)



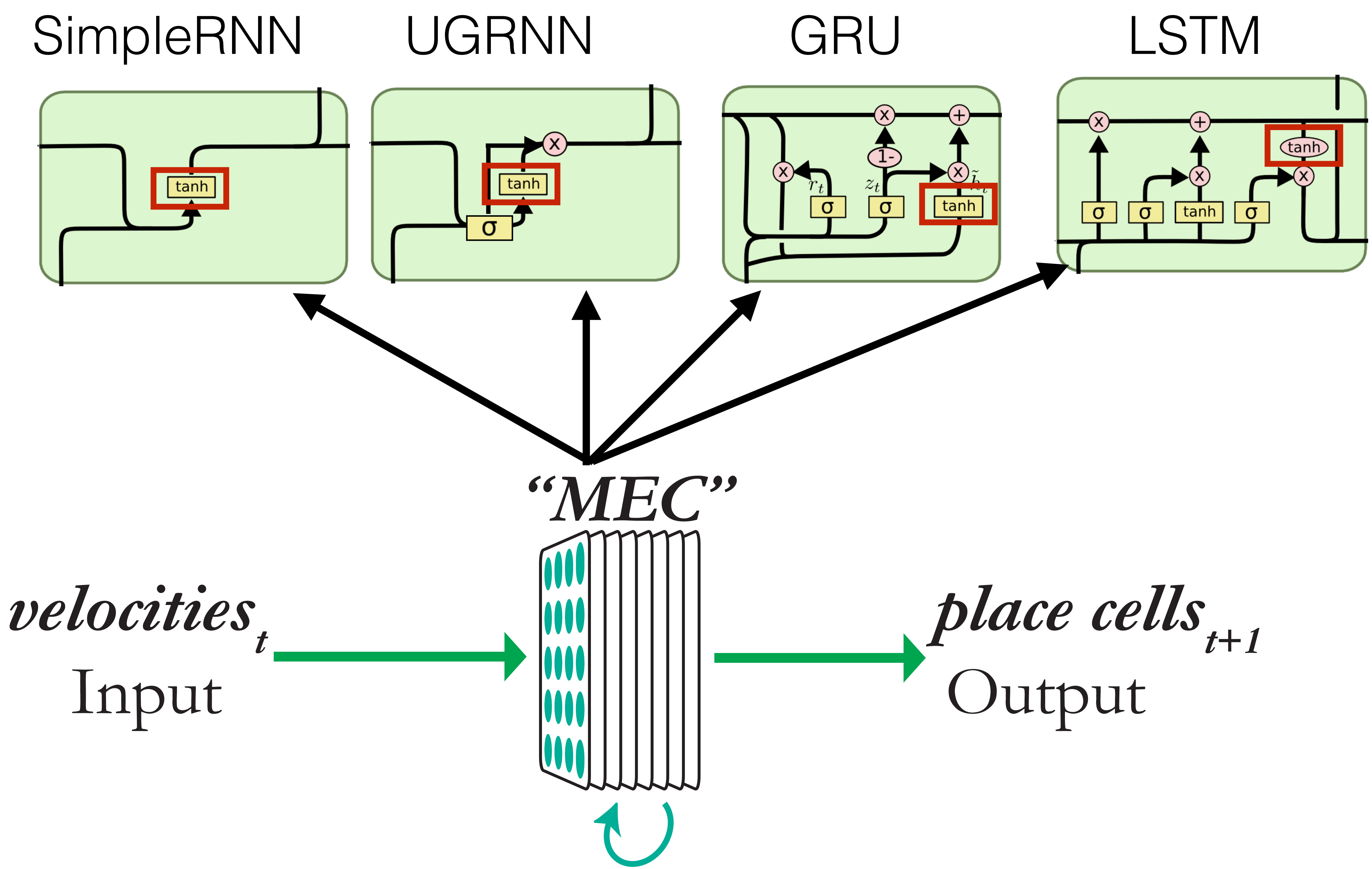
A spectrum of circuits — learnable modulation ("gating")



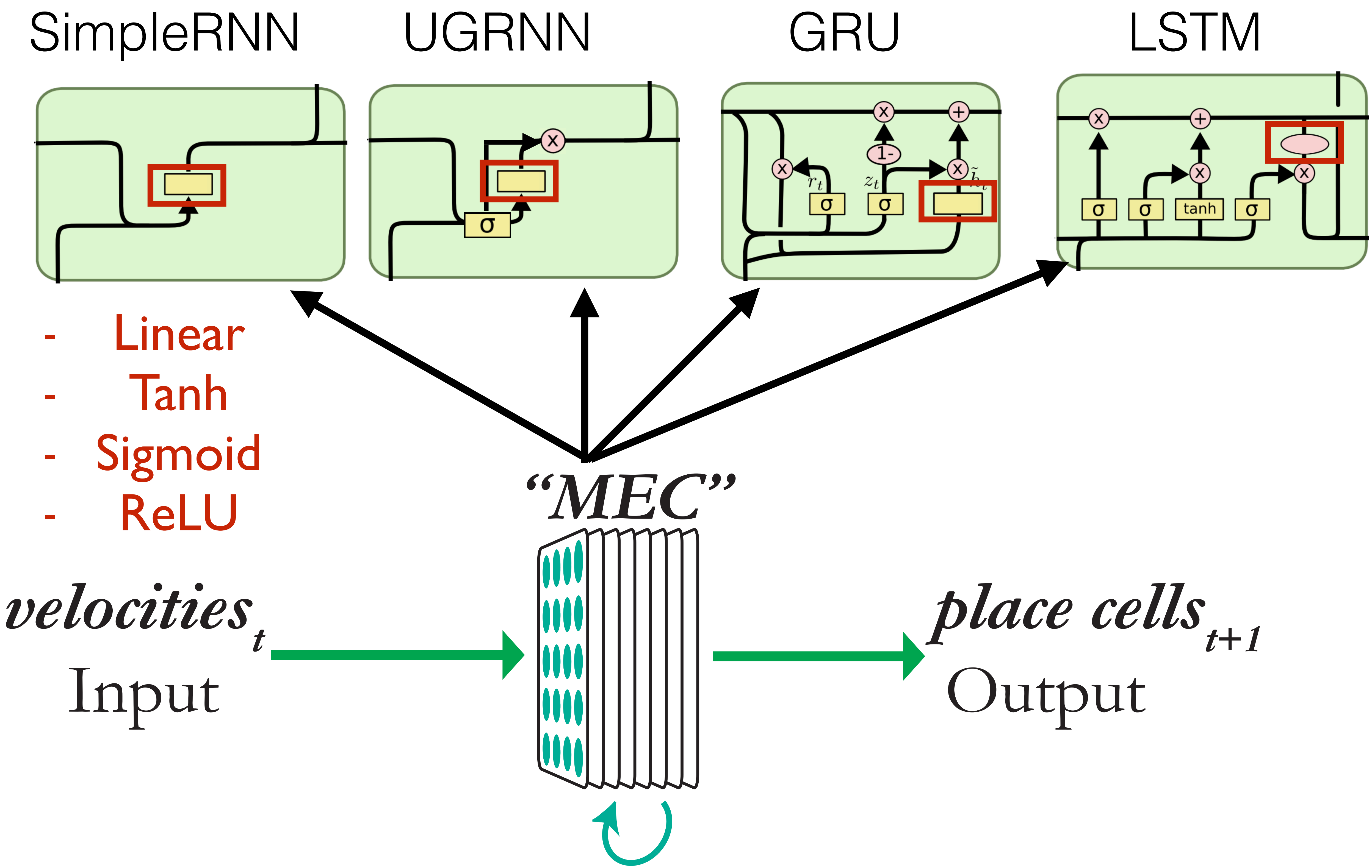
A spectrum of circuits — learnable modulation ("gating")



A spectrum of circuits — output nonlinearity

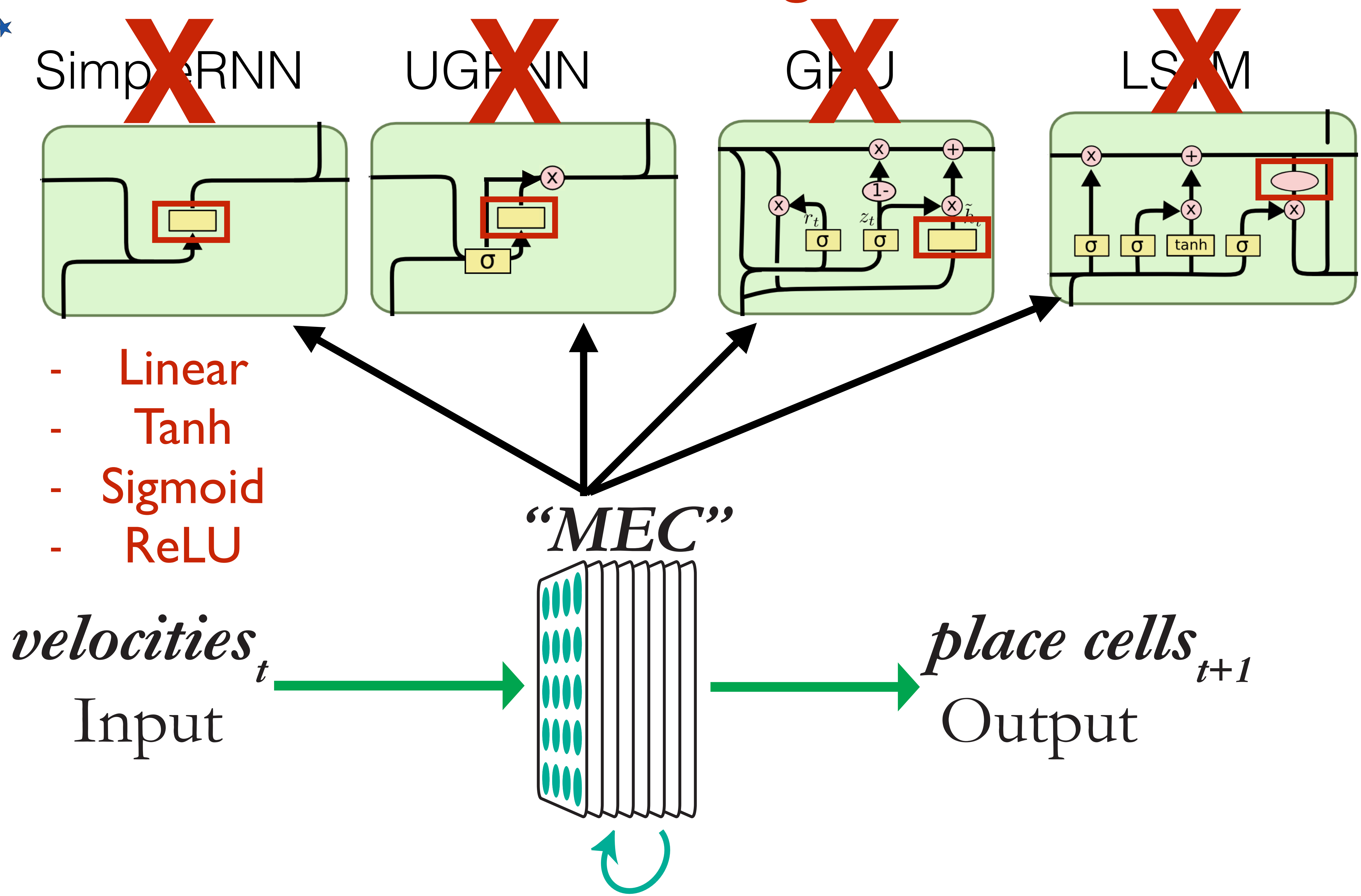


A spectrum of circuits — output nonlinearity

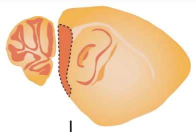


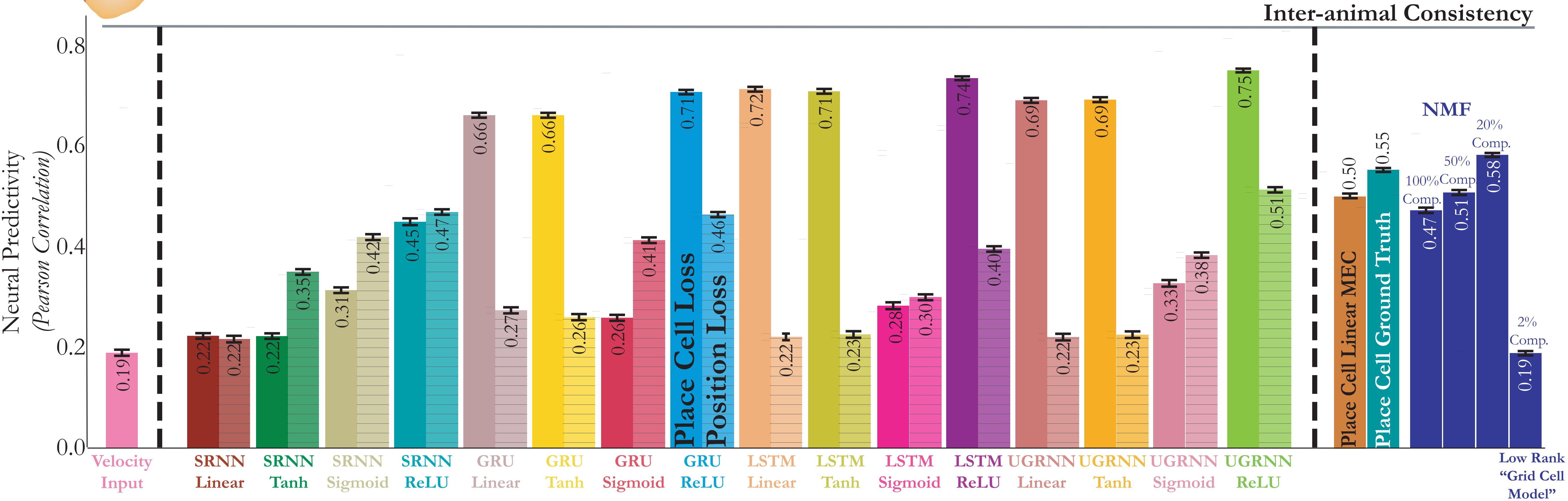
A spectrum of circuits — output nonlinearity

Circuit busting!



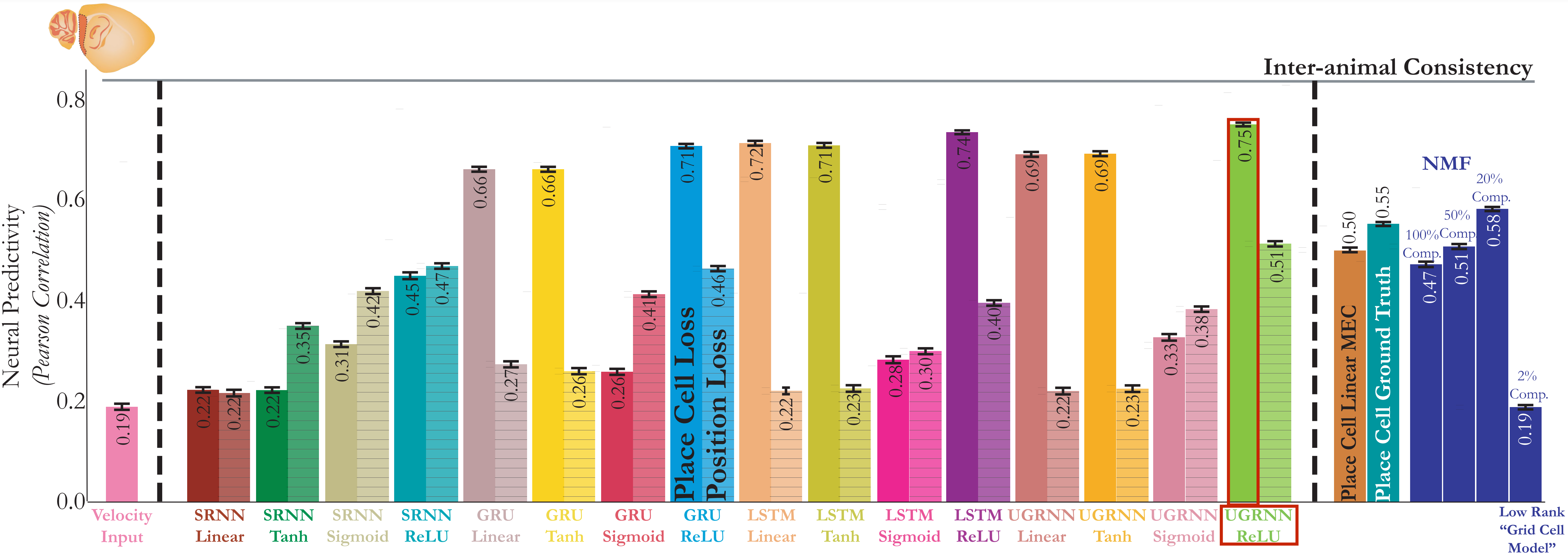
Benchmarking models with the same transform as between animals



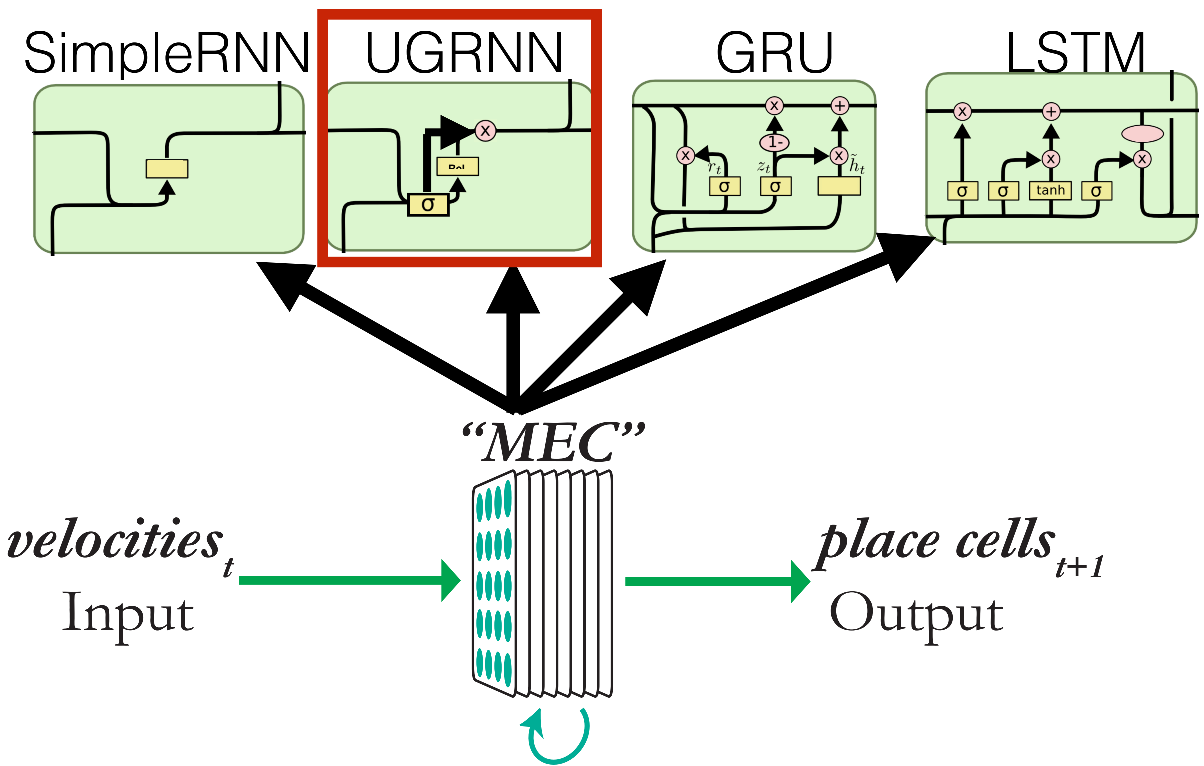


Caitlin Mallory

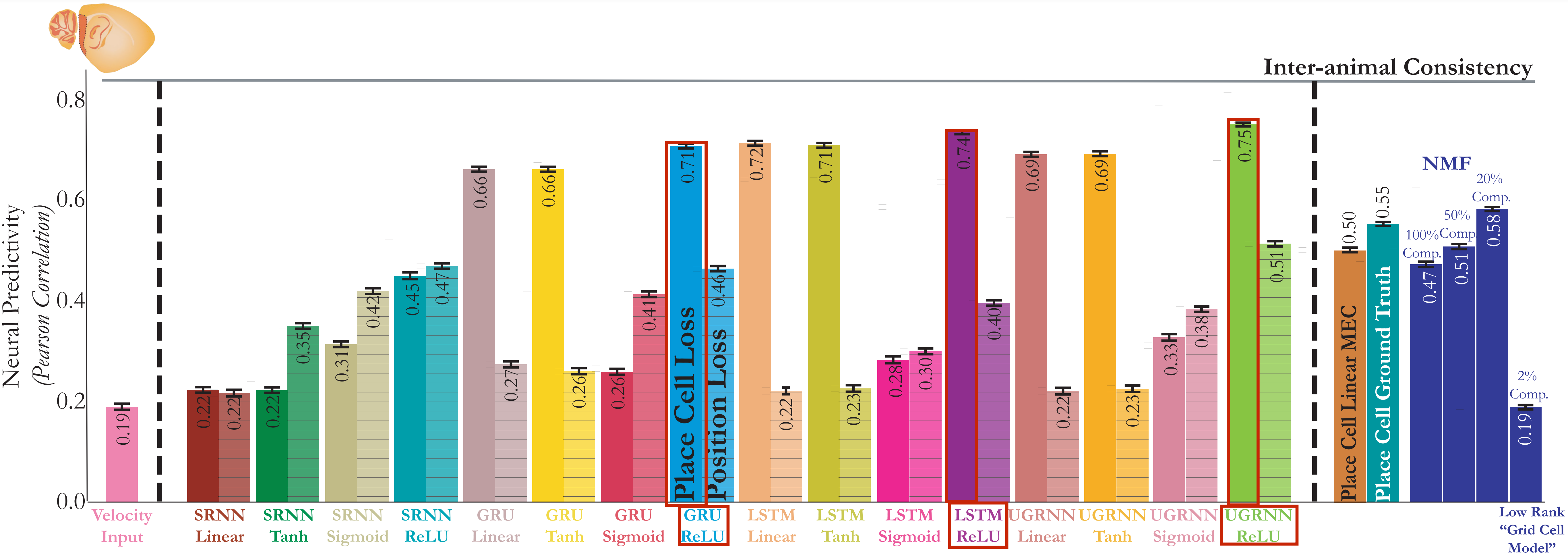
Task-optimized navigational models best predict the *entire* MEC population



Best task-optimized models explain almost all of the neural variability

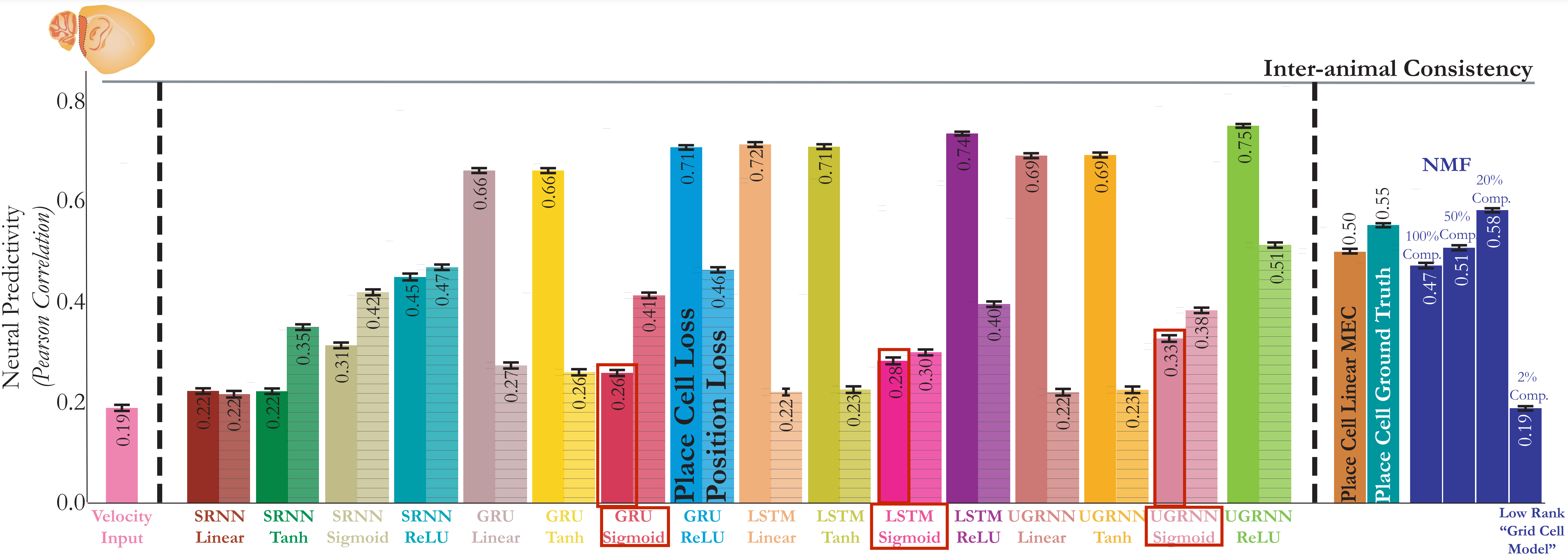


Nonlinearity type affects generalization



Nonnegativity constraint + gating aids in generalization across environments

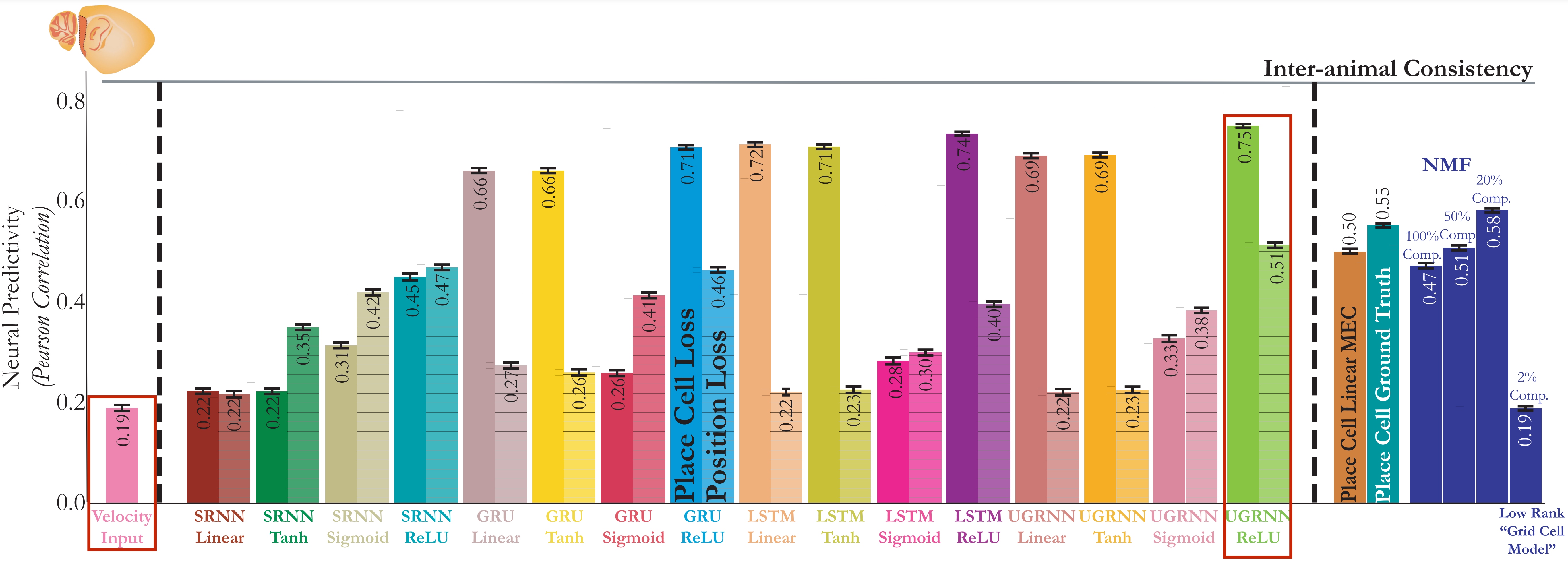
Nonlinearity type affects generalization



Nonnegativity constraint + gating aids in generalization across environments

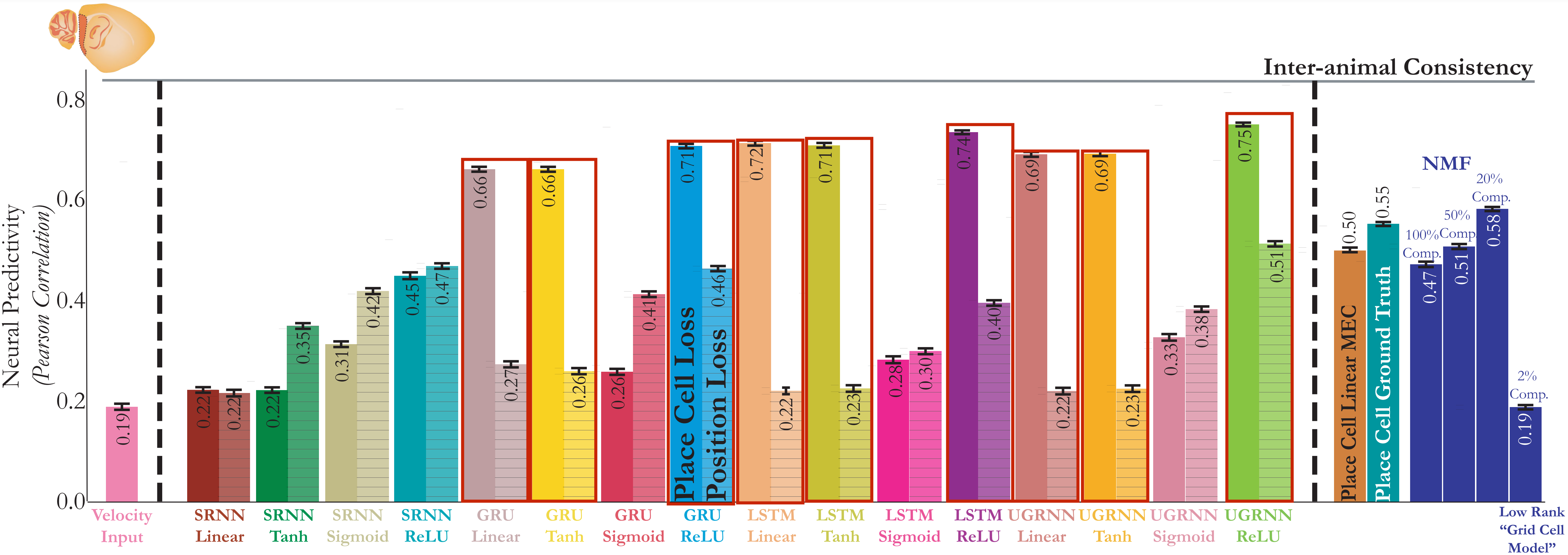
But this nonnegativity constraint must *not* saturate either!

Model input is a poor predictor of population



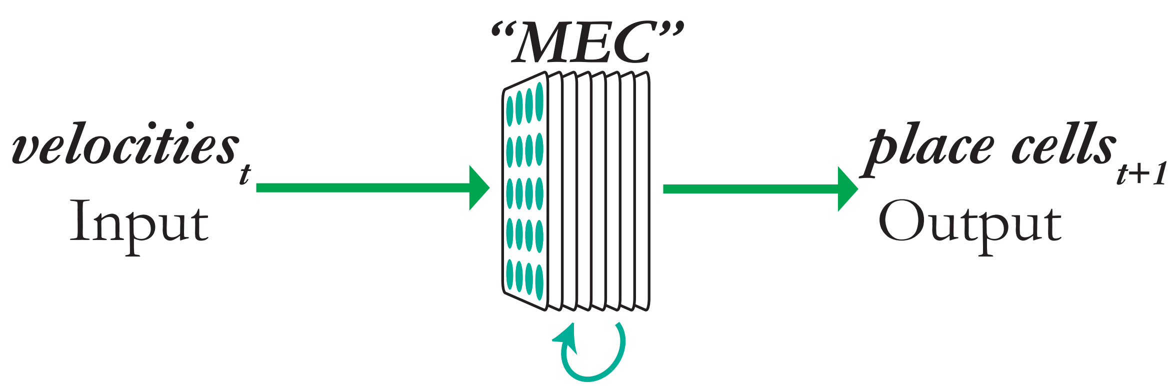
Models add a lot of predictive power to their inputs

Direct path integration *fails* to generalize

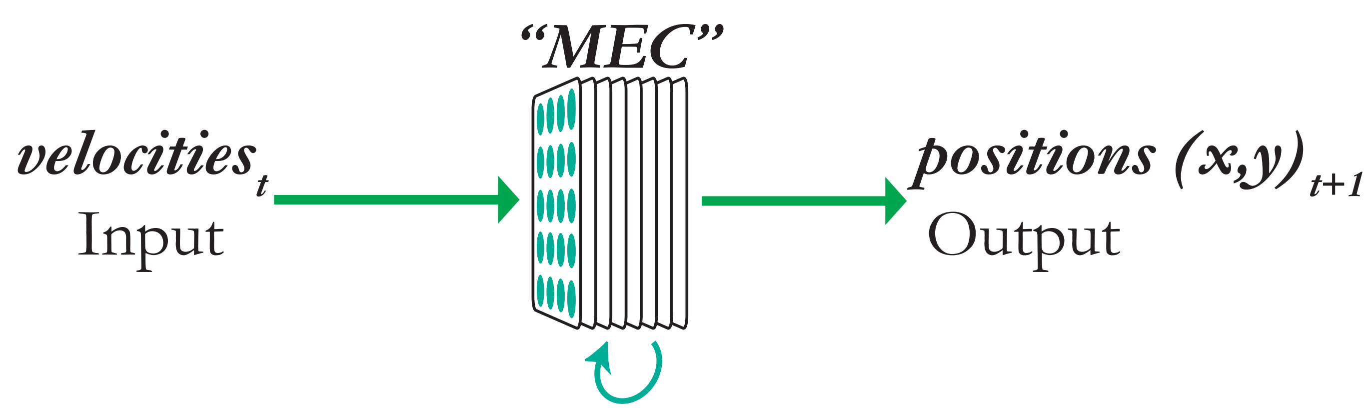


Output place cell supervision provides better generalization over direct supervision of position (path integration)

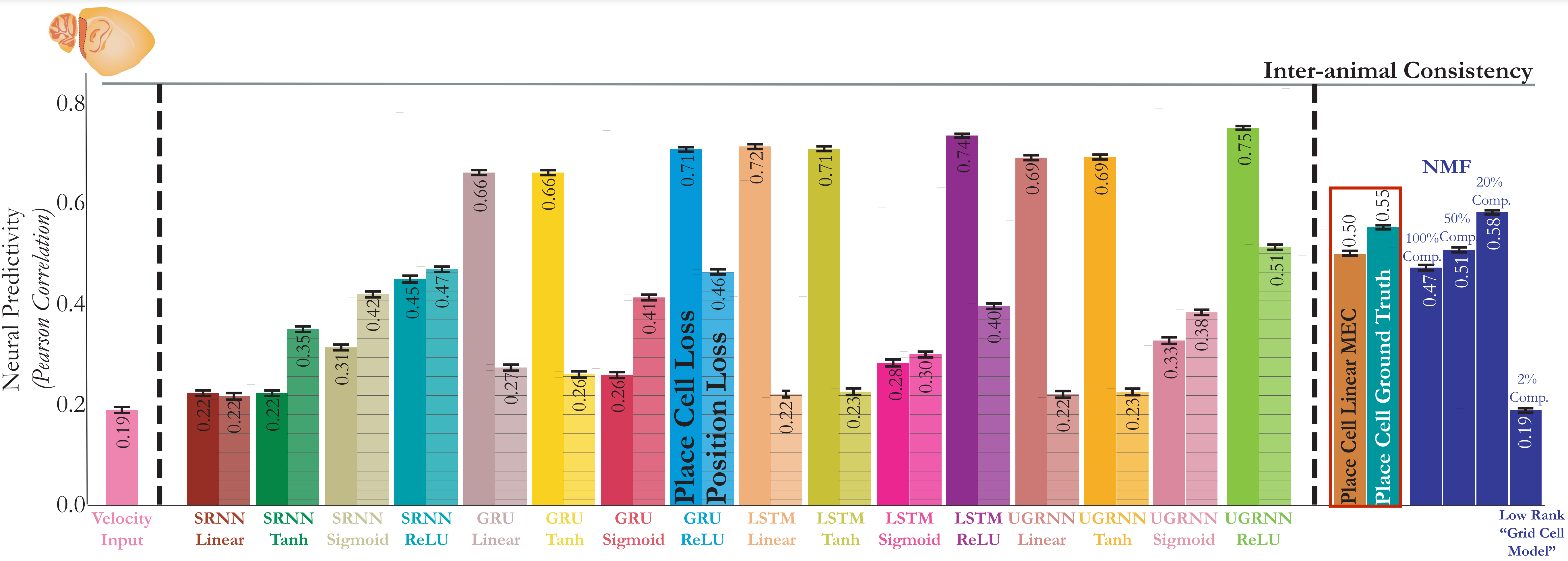
$$\mathcal{L}(\hat{p}, p) := -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N_p} p_i^t \log \hat{p}_i^t$$



$$\mathcal{L}(\hat{p}, p) := \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left(\left(p_x^t - \hat{p}_x^t\right)^2 + \left(p_y^t - \hat{p}_y^t\right)^2\right)$$

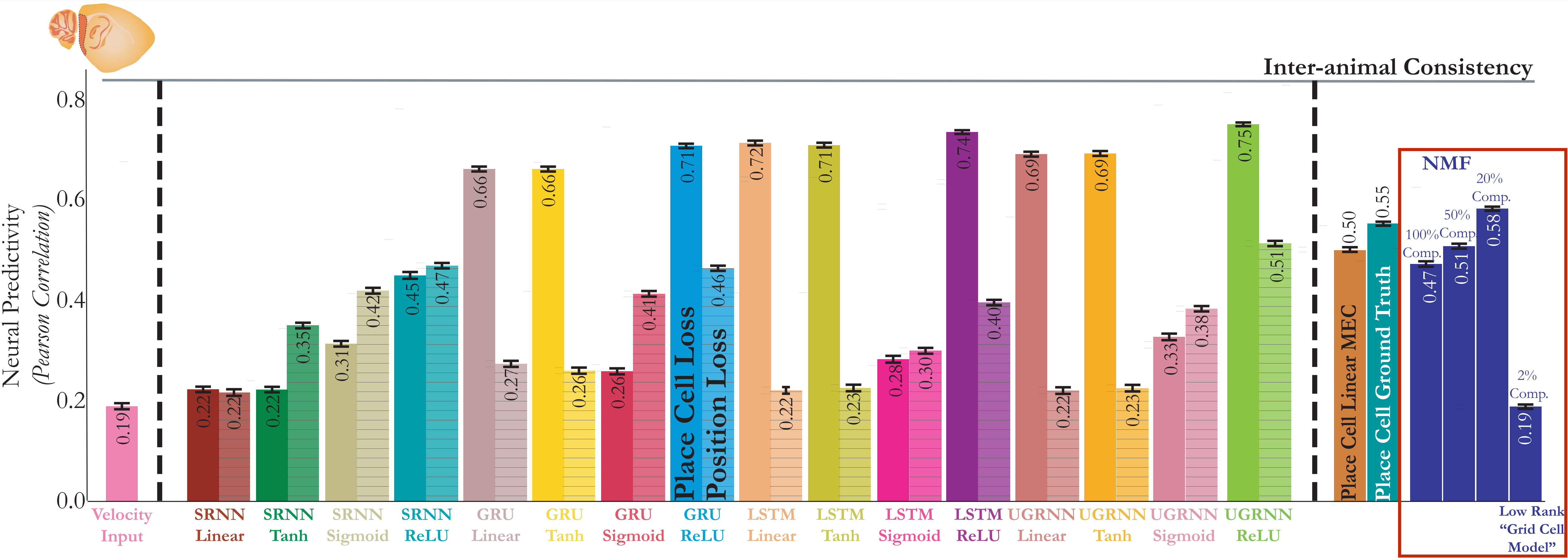


Place cells alone are a poor predictor



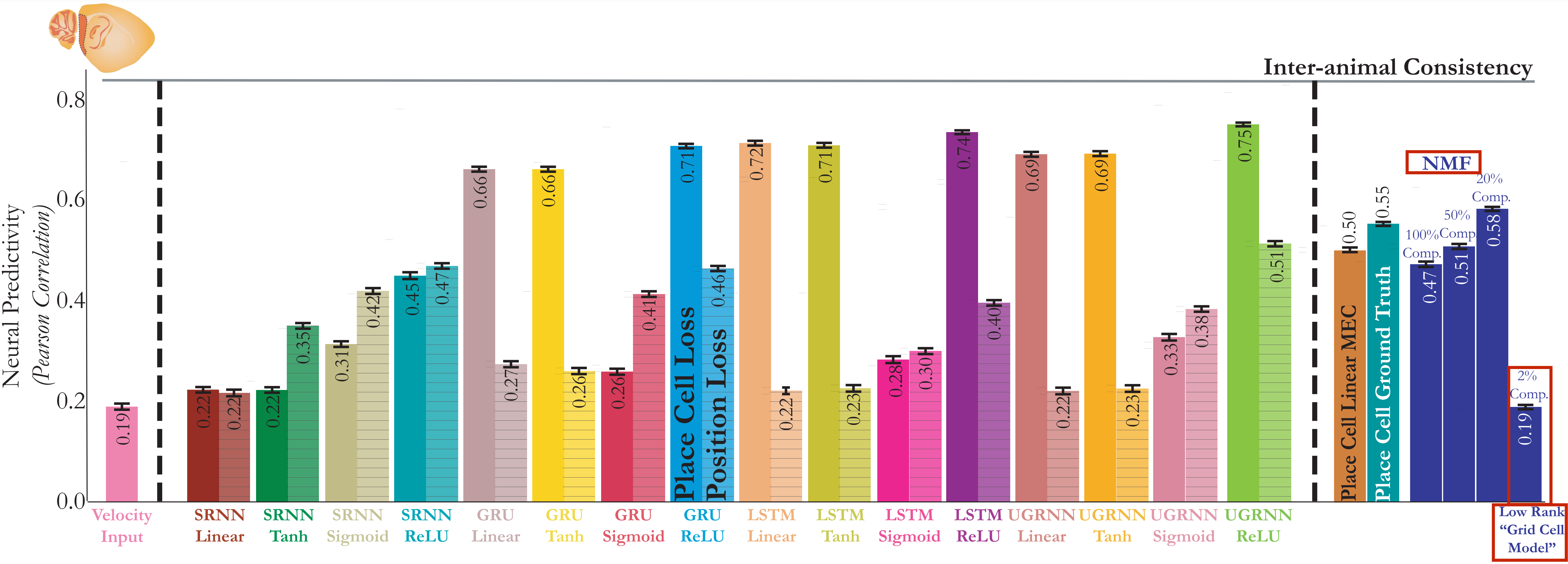
But place cells alone are *not* a good predictor of MEC (good!)
You actually need to integrate them!

NMF is also a poor predictor



Dimensionality reduction on place cells is *not* a good predictor of MEC either

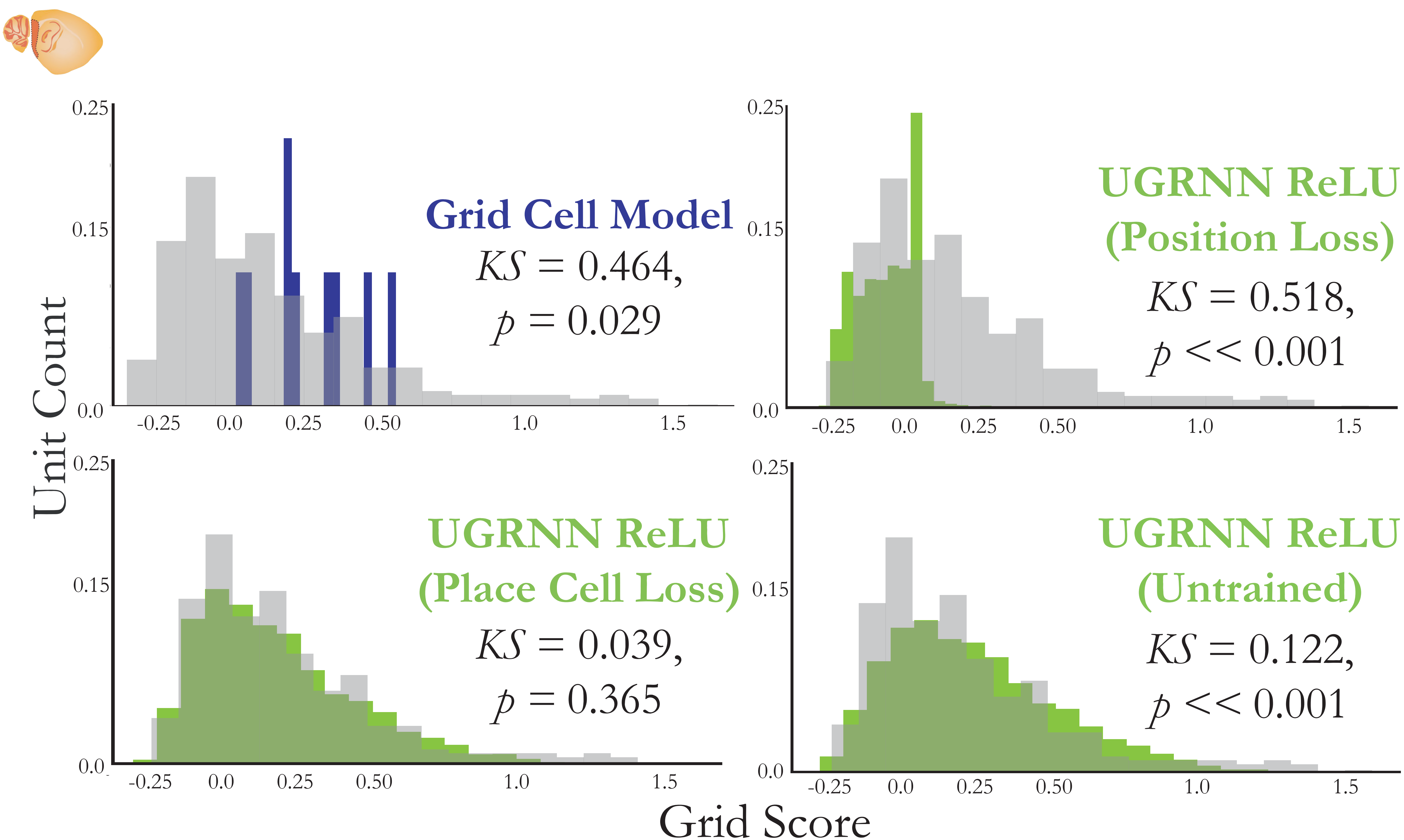
Grid cell oriented NMF is a poor predictor



Grid cell oriented model is an especially *poor* predictor!

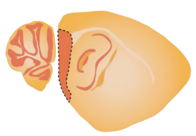
Task-optimized navigational models best predict the *entire* MEC population

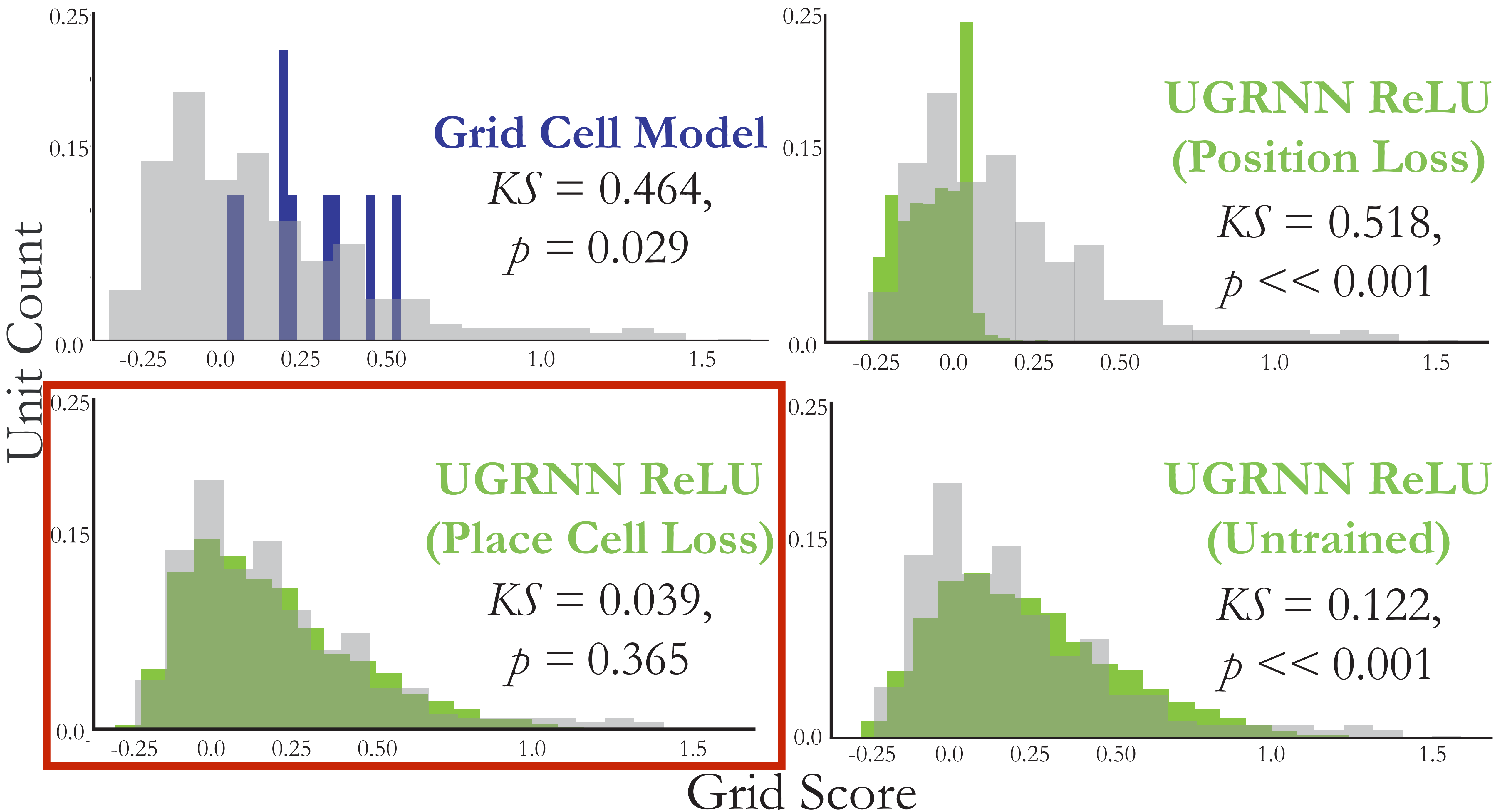
Grid score distribution does not require any parameter fitting



Task-optimized navigational models best predict the *entire* MEC population

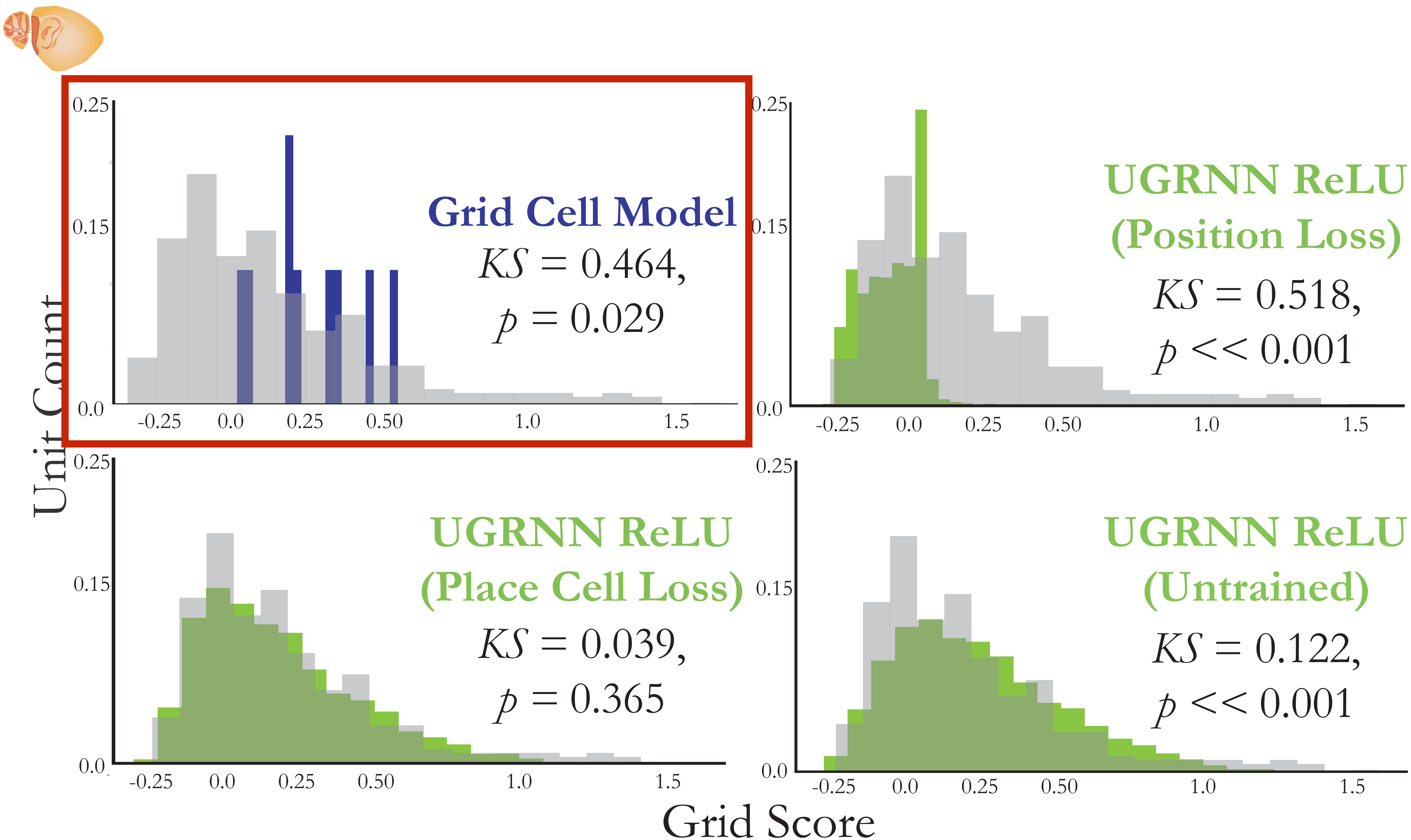
Best model class in terms of neural predictivity also matches grid score distribution in its own synthetic population





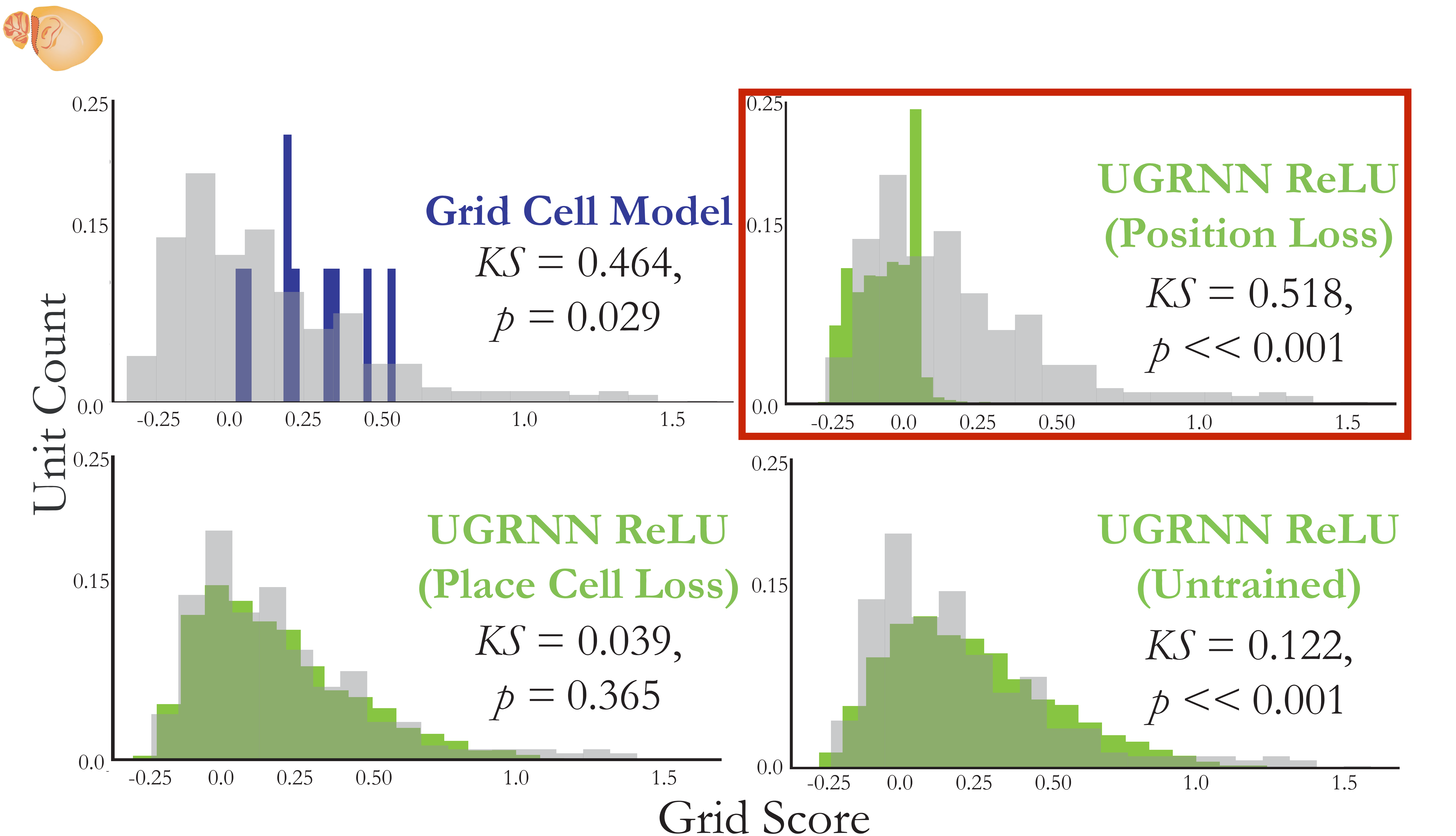
Task-optimized navigational models best predict the *entire* MEC population

Low-rank model is too biased towards grid-like units



Task-optimized navigational models best predict the *entire* MEC population

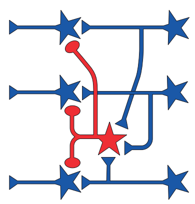
Without place cell integration, the model is too biased towards *non* grid-like units



Takeaways

A = *architecture class*

1. *"Circuit"*



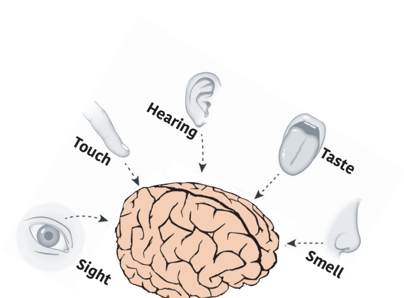
T = *task loss*

3. *"Ecological niche/behavior"*

<u>Neurobiological Puzzle(s):</u>

1. How might we characterize what these heterogeneous cells do?

2. What functional role do these cells serve in the circuit, if any?



2. *"Environment"*

D = *data stream*

Takeaways

A = *architecture class*

1. *"Circuit"*

gating + nonnegativity

T = *task loss*

3. *"Ecological niche/behavior"*

place cell integration

~~path integration~~

<u>Neurobiological Puzzle(s)</u>:

1. How might we characterize what these heterogeneous cells do?

2. What functional role do these cells serve in the circuit, if any?

<u>*Partial* Resolution:</u>

1. <u>Characterization:</u> Close to perfect neural predictivity with the above constraints — more complex environments are needed!

2. <u>Functional Role:</u> Grid cells are not functionally unique! Both heterogeneous and grid cells arise jointly through task optimization.

2. *"Environment"*

D = *data stream*

Broad Takeaways

- Recurrent Connections in the Primate Ventral Stream

Enable a high performing network to fit in cortex, attaining computational power through temporal rather than spatial complexity during core object recognition.

- Goal-Driven Models of Mouse Visual Cortex

Low-resolution, shallow network that makes best use of the mouse's limited resources to create a light-weight, general-purpose visual system.

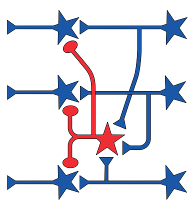
- Heterogeneity in Rodent Medial Entorhinal Cortex

Heterogeneous cells are <u>not</u> functionally segregated from classic cell types, but rather form a continuum of cells, shaped by a process of biological performance optimization.

Building and Identifying Biologically-Plausible Learning Rules

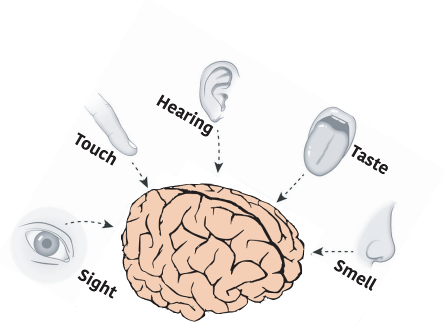
A = *architecture class*

1. *"Circuit"*



T = *task loss*

3. *"Ecological niche/behavior"*



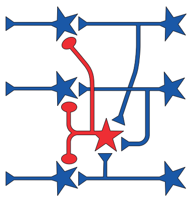
2. *"Environment"*

D = *data stream*

Building and Identifying Biologically-Plausible Learning Rules

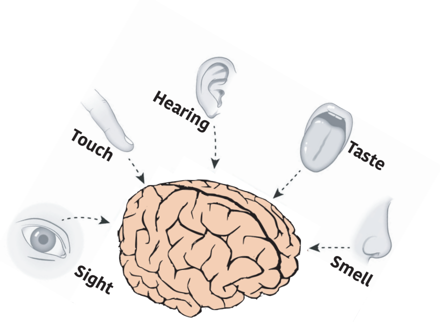
A = *architecture class*

1. *"Circuit"*



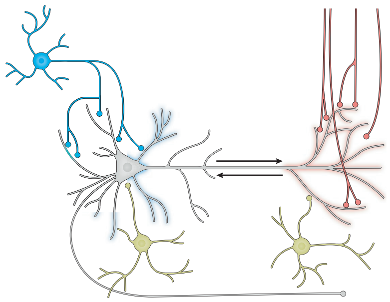
T = *task loss*

3. *"Ecological niche/behavior"*



2. *"Environment"*

D = *data stream*



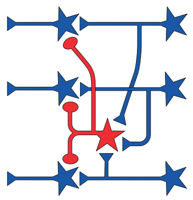
4. *"Developmental mechanism"*

Learning Rule

Building and Identifying Biologically-Plausible Learning Rules

A = *architecture class*

1. *"Circuit"*



T = *task loss*

3. *"Ecological niche/behavior"*

Goal: Identifying Learning Rules in Neural Circuits


Dan Kunin


Javier Sagastuy-Brena


Jon Bloom

Two Routes to Scalable Credit Assignment without Weight Symmetry

Int'l Conf. Mach. Learn. 2020

* Making local rules that actually *work* at scale
* "Localizing" apparently non-local rules


Surya Ganguli


Sanjana Srivastava

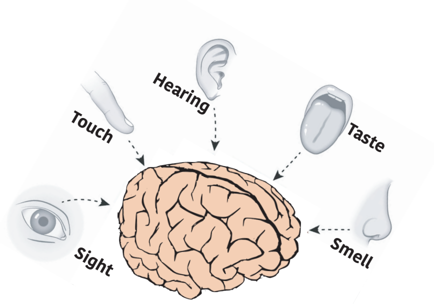
Identifying Learning Rules from Neural Network Observables

NeurIPS 2020 (chosen for spotlight presentation)


Daniel Yamins

* How would you even tell if a given rule was at work from actual neural data?
* Might actually be possible to tell from activities alone!

(Hard to record from synapses in vivo)



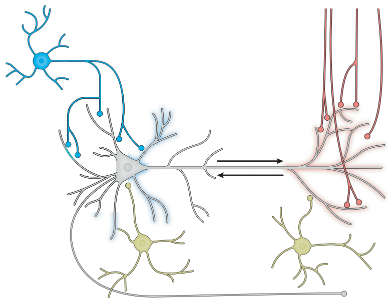
2. *"Environment"*

D = *data stream*


Joshua Melander

Distinct in vivo dynamics of excitatory synapses onto cortical pyramidal neurons and parvalbumin-positive interneurons

Cell Reports 2021



4. *"Developmental mechanism"*

Learning Rule

Outline

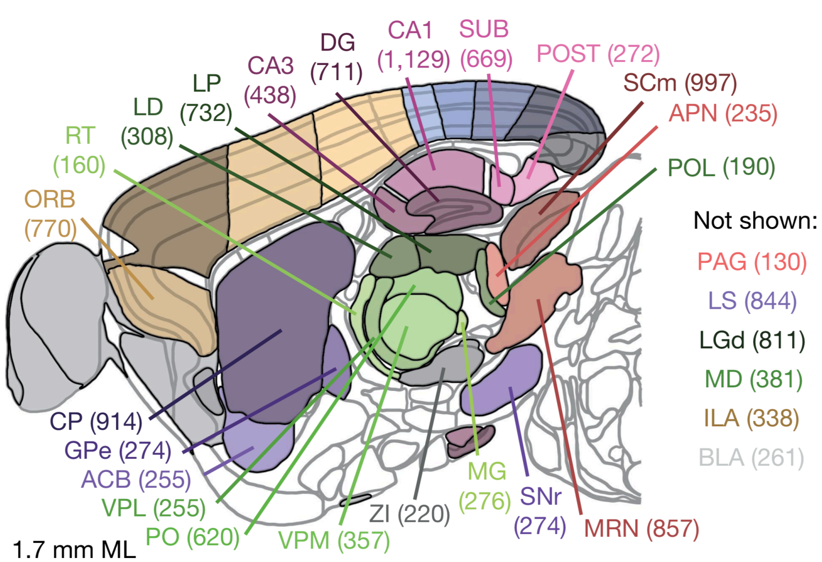
- Recurrent Connections in the Primate Ventral Stream
- Goal-Driven Models of Mouse Visual Cortex
- Heterogeneity in Rodent Medial Entorhinal Cortex
- Future Directions

Next Steps

Next Steps

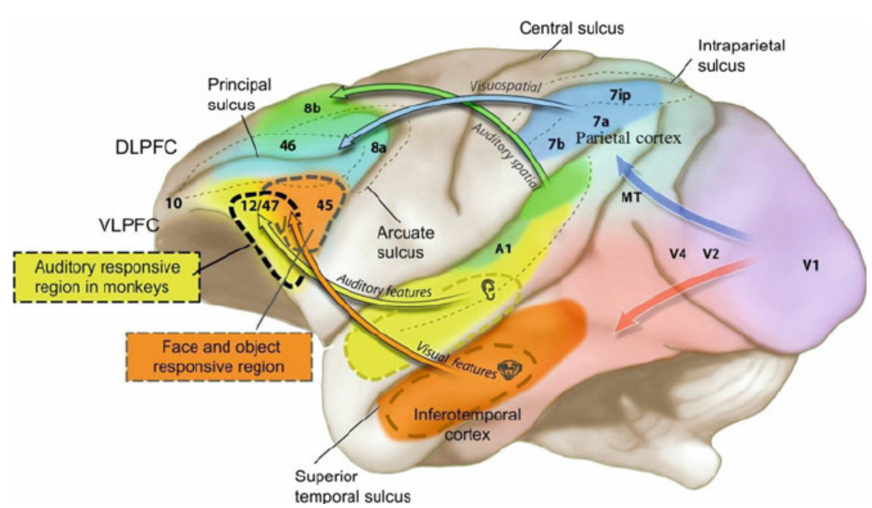
Decades of neuroscience research have revealed that the brain has a modular, integrative design that involves the coordination of multiple subsystems.

Mouse



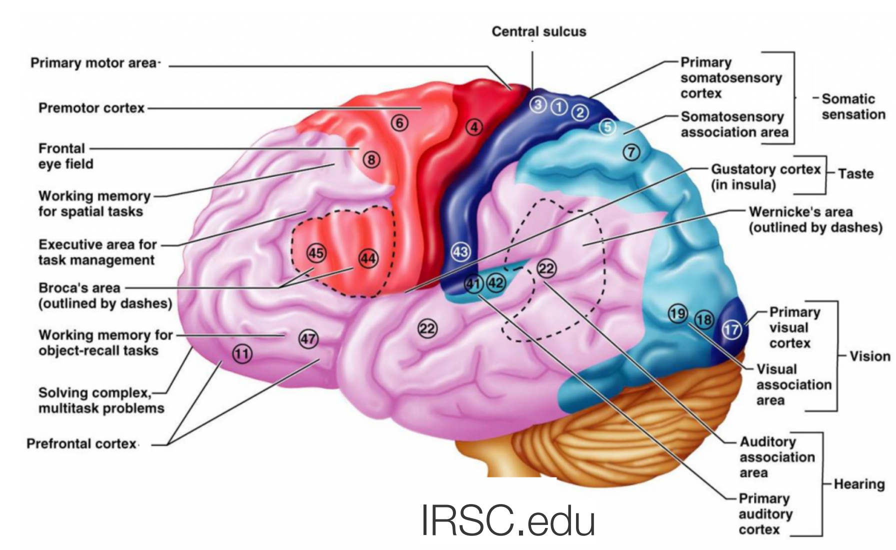
Steinmetz, Zatka-Haas, Carandini, Harris *Nature* 2019

Macaque



Poon *et al.* 2013

Human



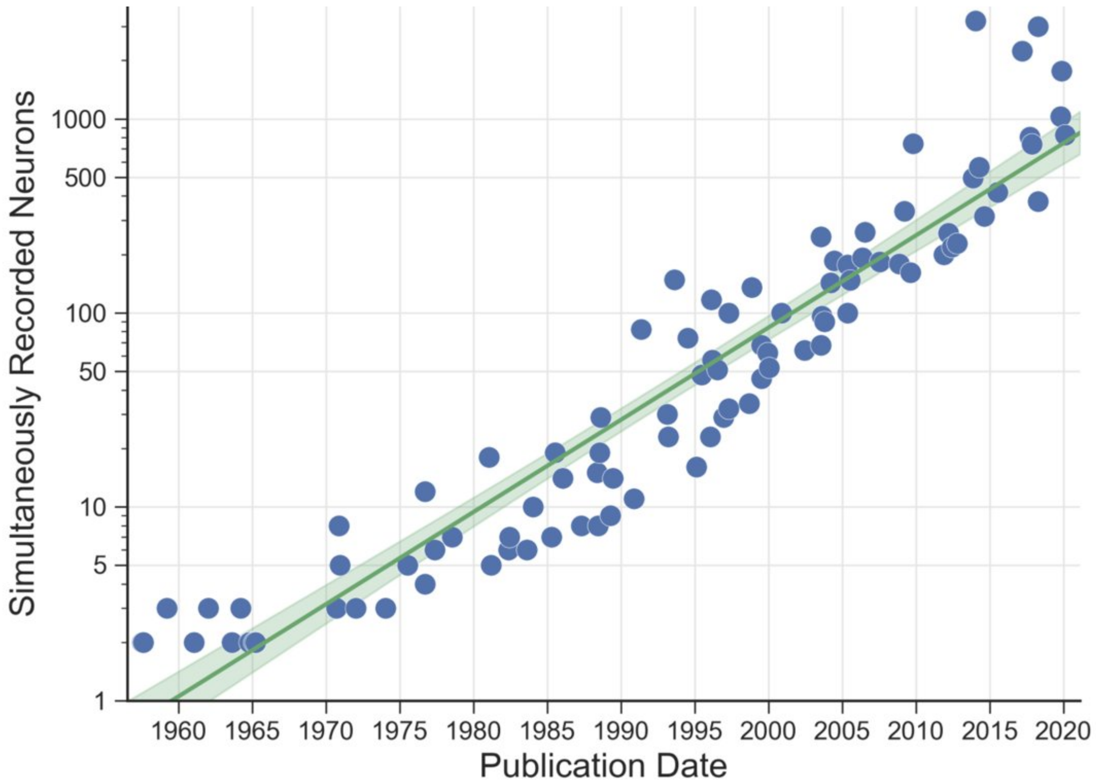
IRSC.edu

Next Steps

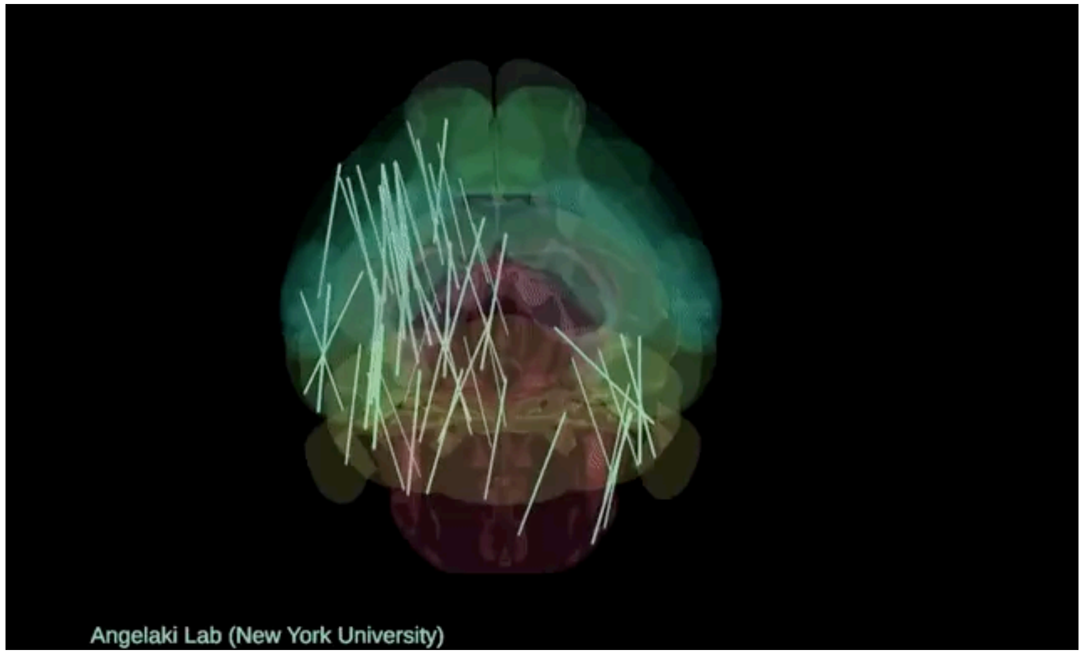
Decades of neuroscience research have revealed that the brain has a modular, integrative design that involves the coordination of multiple subsystems.

Large-scale neural data are increasingly collected in awake, behaving animals — our understanding of these data necessitates that we begin to understand how these systems combine together.

Number of neurons recorded doubling every 6 years



Ian Stevenson



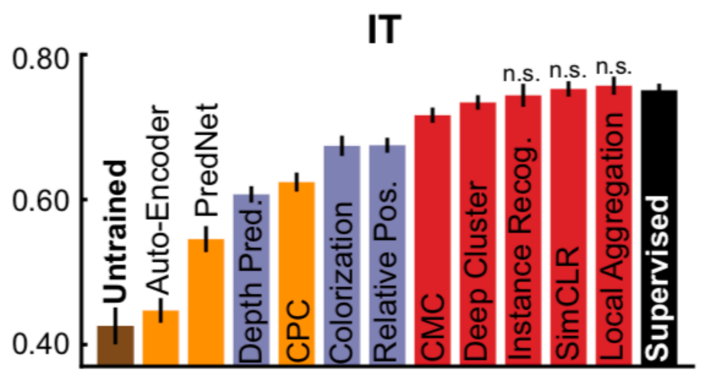
International Brain Laboratory 2022

Next Steps

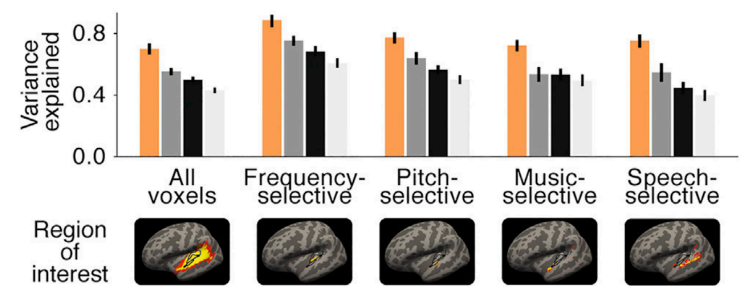
Goal-driven models have typically been functionally restricted to a single system, and are structurally mapped to one neuroanatomical pathway.

<u>Reasonable models for individual systems</u>

Visual System



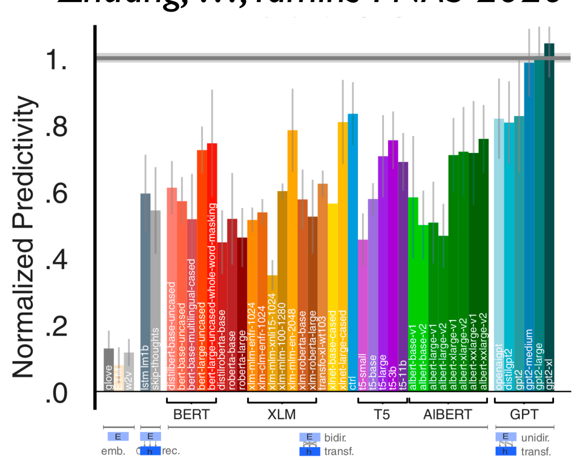
Zhuang, …, Yamins PNAS 2020



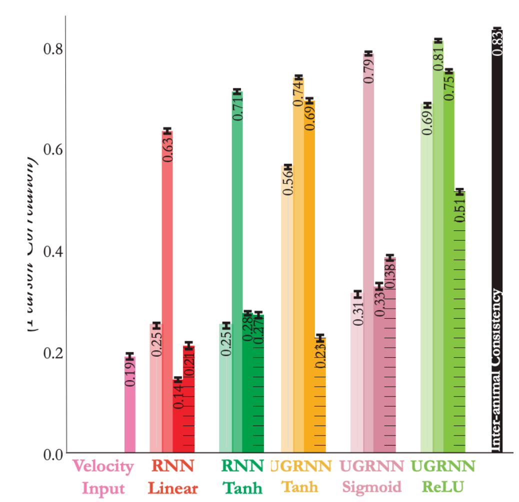
Auditory System

Kell, Yamins* …, McDermott Neuron 2018*

Language System



Schrimpf, …, Fedorenko PNAS 2021



Entorhinal-Hippocampal System

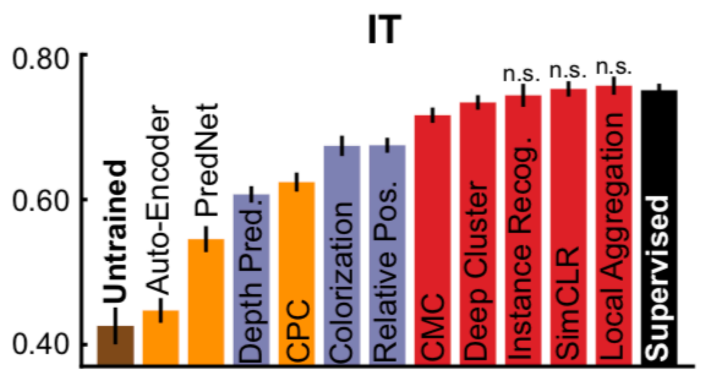
***Nayebi,** …, Yamins NeurIPS 2021*

Next Steps

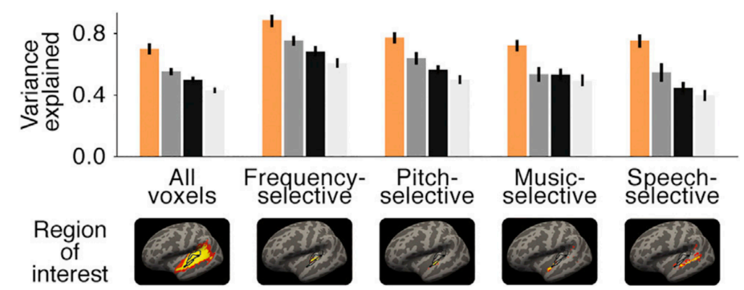
Goal-driven models have typically been functionally restricted to a single system, and are structurally mapped to one neuroanatomical pathway.

But these systems do not generate all behaviors on their own!

Visual System



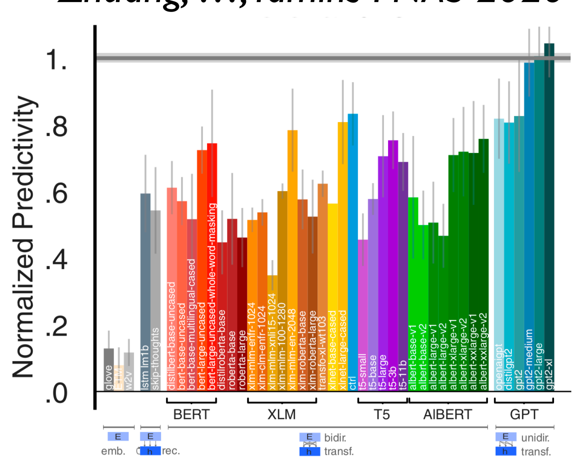
Zhuang, …, Yamins PNAS 2020



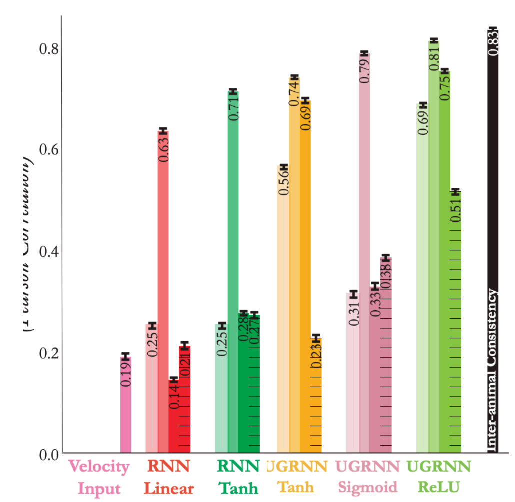
Auditory System

Kell, Yamins* …, McDermott Neuron 2018*

Language System



Schrimpf, …, Fedorenko PNAS 2021



Entorhinal-Hippocampal System

***Nayebi,** …, Yamins NeurIPS 2021*

Complex Behaviors *Require* Integration

Multi-modal perception
(e.g. vision, audition, olfaction, somatosensory cortices)

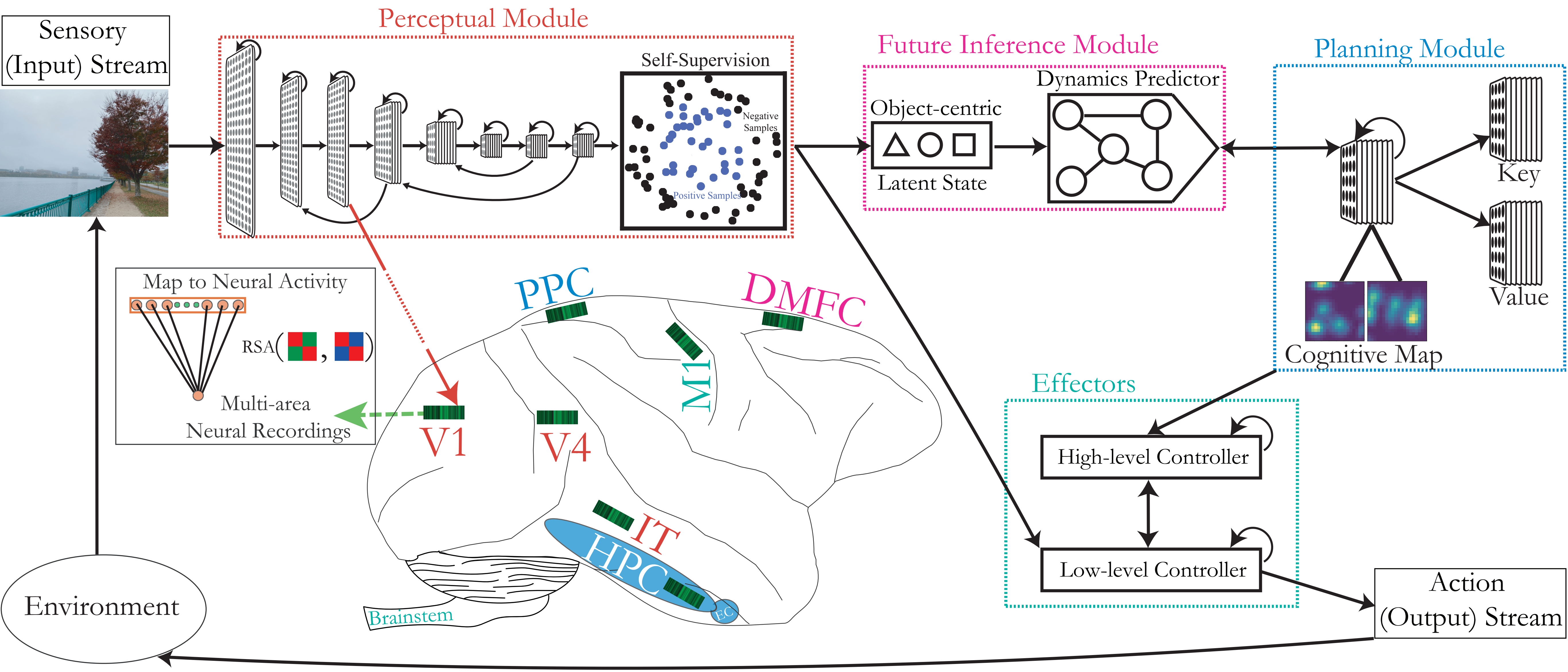
Object manipulation
(requires perception, intuitive physics, and motor coordination — fronto-parietal network, visual, and motor cortices)

Autonomous navigation in challenging environments
(coordination of e.g. visual, entorhinal, hippocampus, and motor cortices)

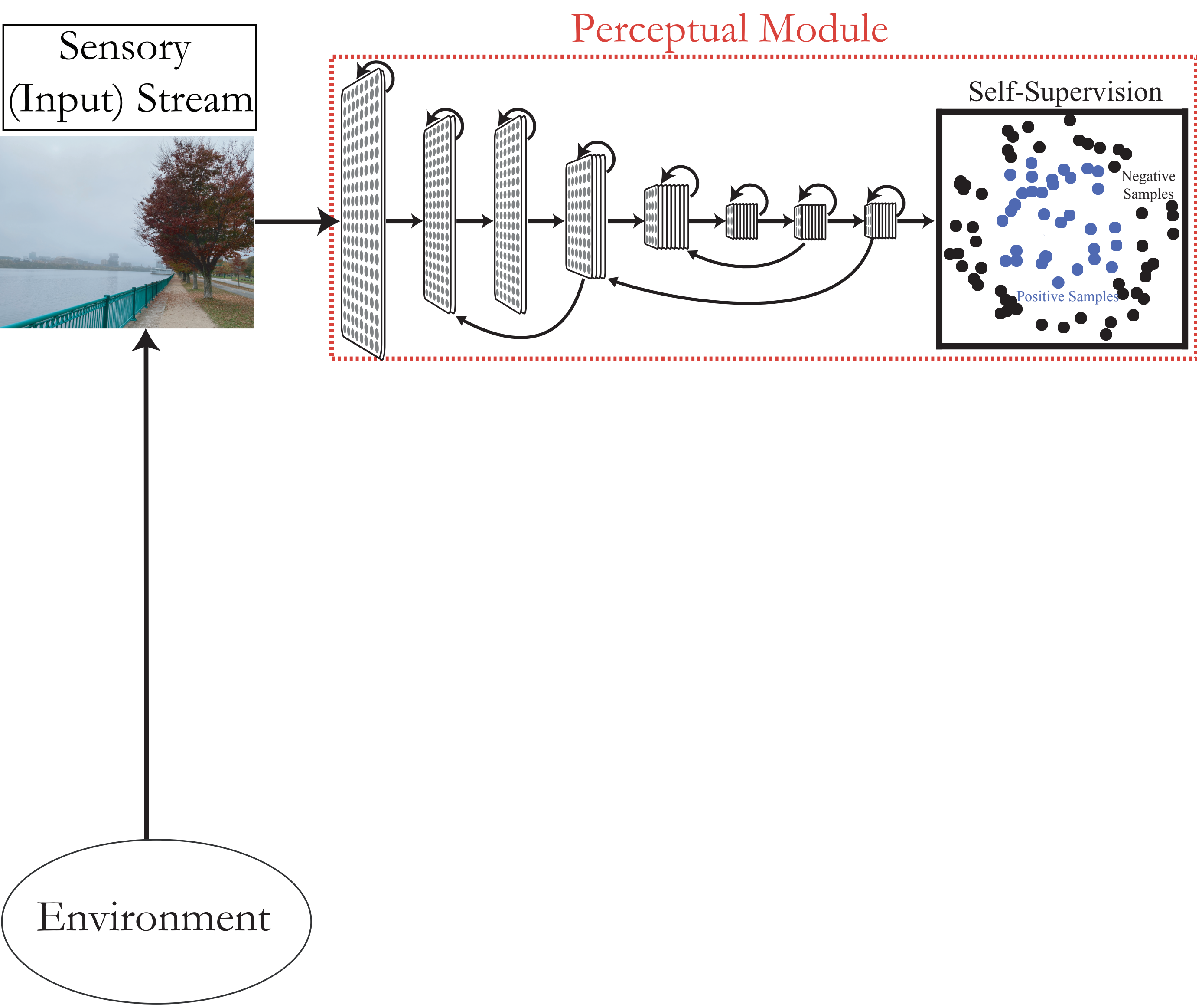
Sensorimotor coordination in complex, unpredictable environments
(basal ganglia, brainstem, motor cortices)

Relational inference and planning
(coordination of e.g. sensory, hippocampus, and parietal cortices)

Next Steps: Integrative Agents to Reverse-Engineer Natural Cognition



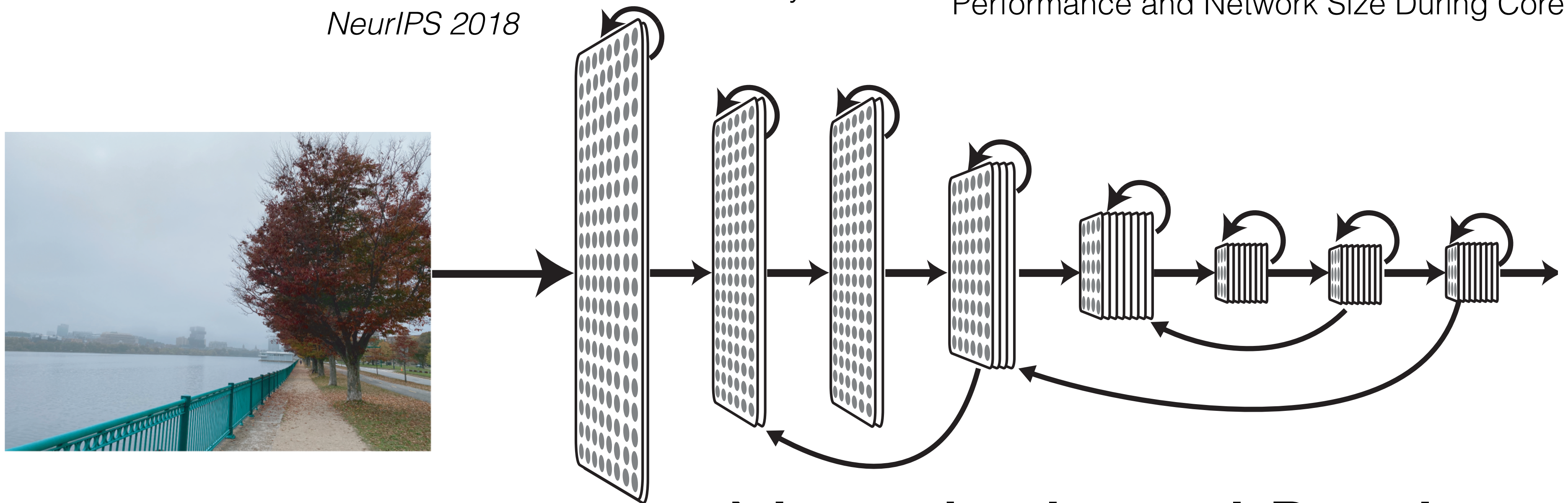
Next Steps: Improving Perceptual Intelligence



Next Steps: Improving Perceptual Intelligence

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*



<u>Neurobiological Puzzle:</u>

What is the role of recurrence in the primate ventral stream?

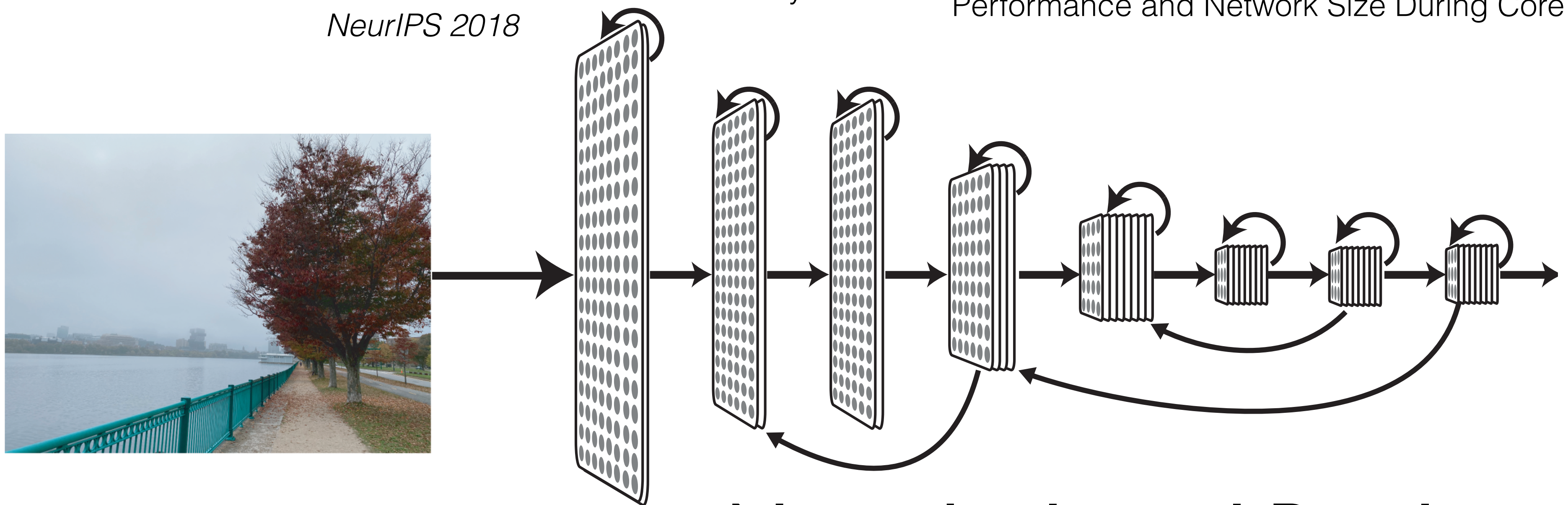
<u>*Partial* Resolution:</u>

Enables high performance by trading off space with time,
in particular space ~ # of neurons (<u>not</u> # of synapses).

Next Steps: Improving Perceptual Intelligence

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*



<u>Neurobiological Puzzle:</u>

What is the role of recurrence in the primate ventral stream?

<u>*Partial* Resolution:</u>

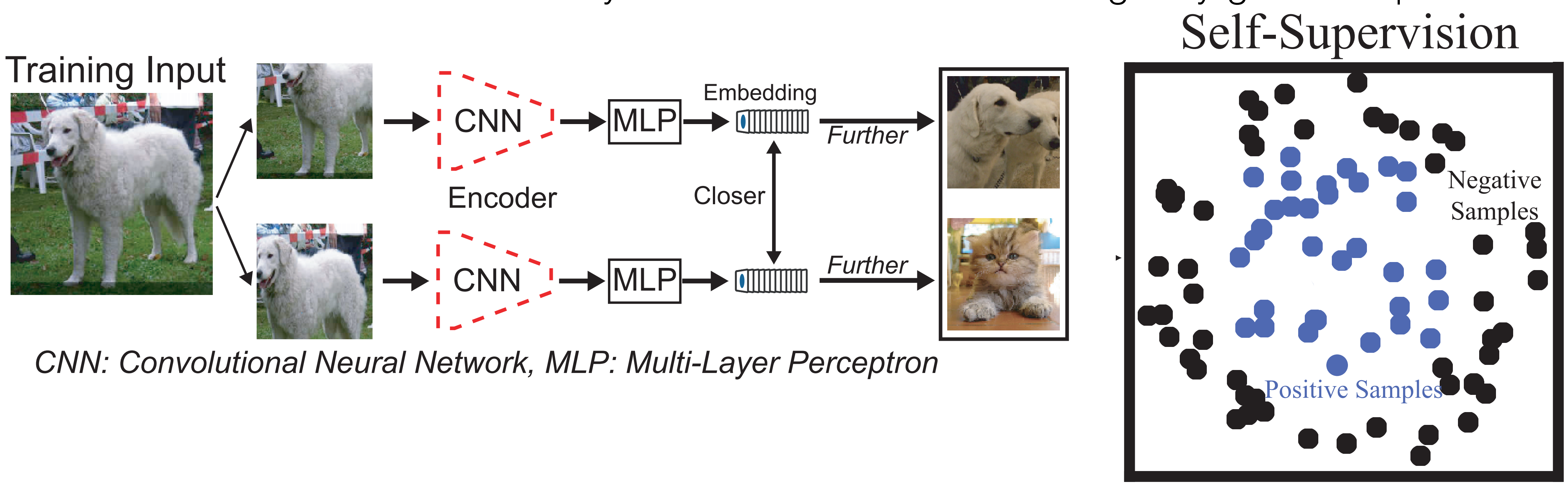
Enables high performance by trading off space with time, *in particular* space ~ # of neurons (<u>not</u> # of synapses).

<u>*Core Unsolved Question:*</u>

What about in <u>dynamic</u> scenes beyond the first 200 *ms*?

Next Steps: Improving Perceptual Intelligence

A. Nayebi*, N. Kong* *et al.*
Mouse visual cortex as a limited resource system that self-learns an ecologically-general representation. *bioRxiv (2021)*



CNN: Convolutional Neural Network, MLP: Multi-Layer Perceptron

<u>Neurobiological Puzzle:</u>

What are the core differences between mouse visual cortex and the primate ventral stream?

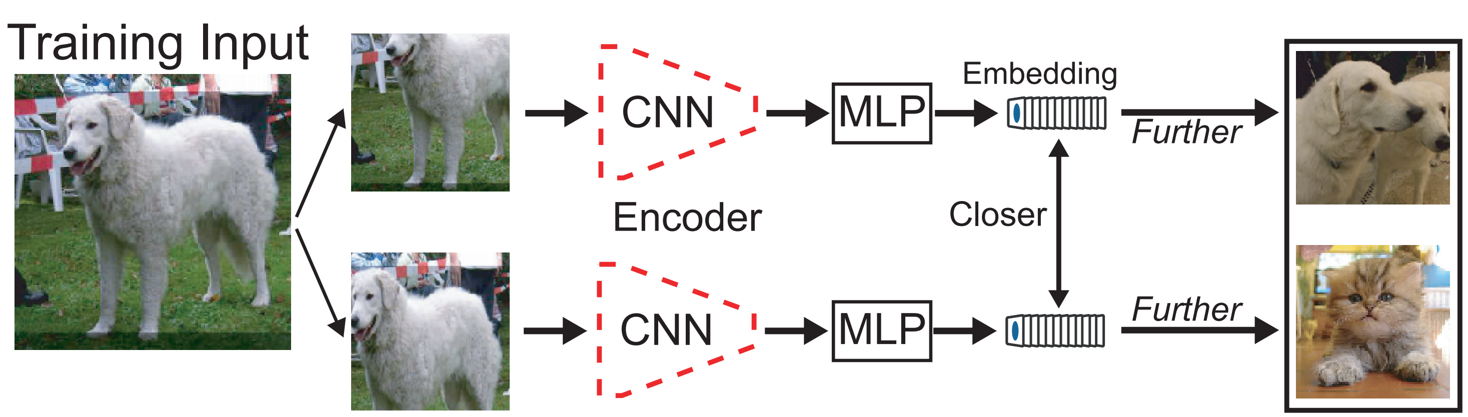
<u>*Partial* Resolution:</u>

Mouse visual cortex is a general-purpose system utilizing its limited resources to perform a variety of visual tasks.

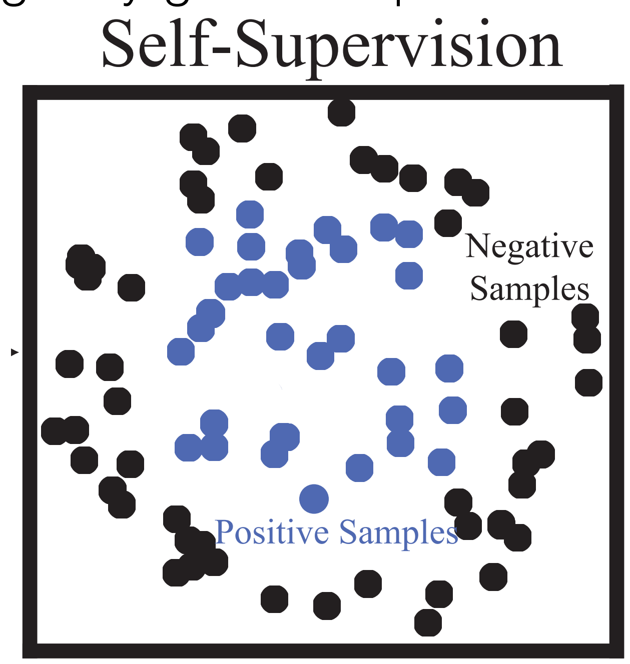
In contrast to the deep, high-resolution, and *task-specific* primate ventral stream.

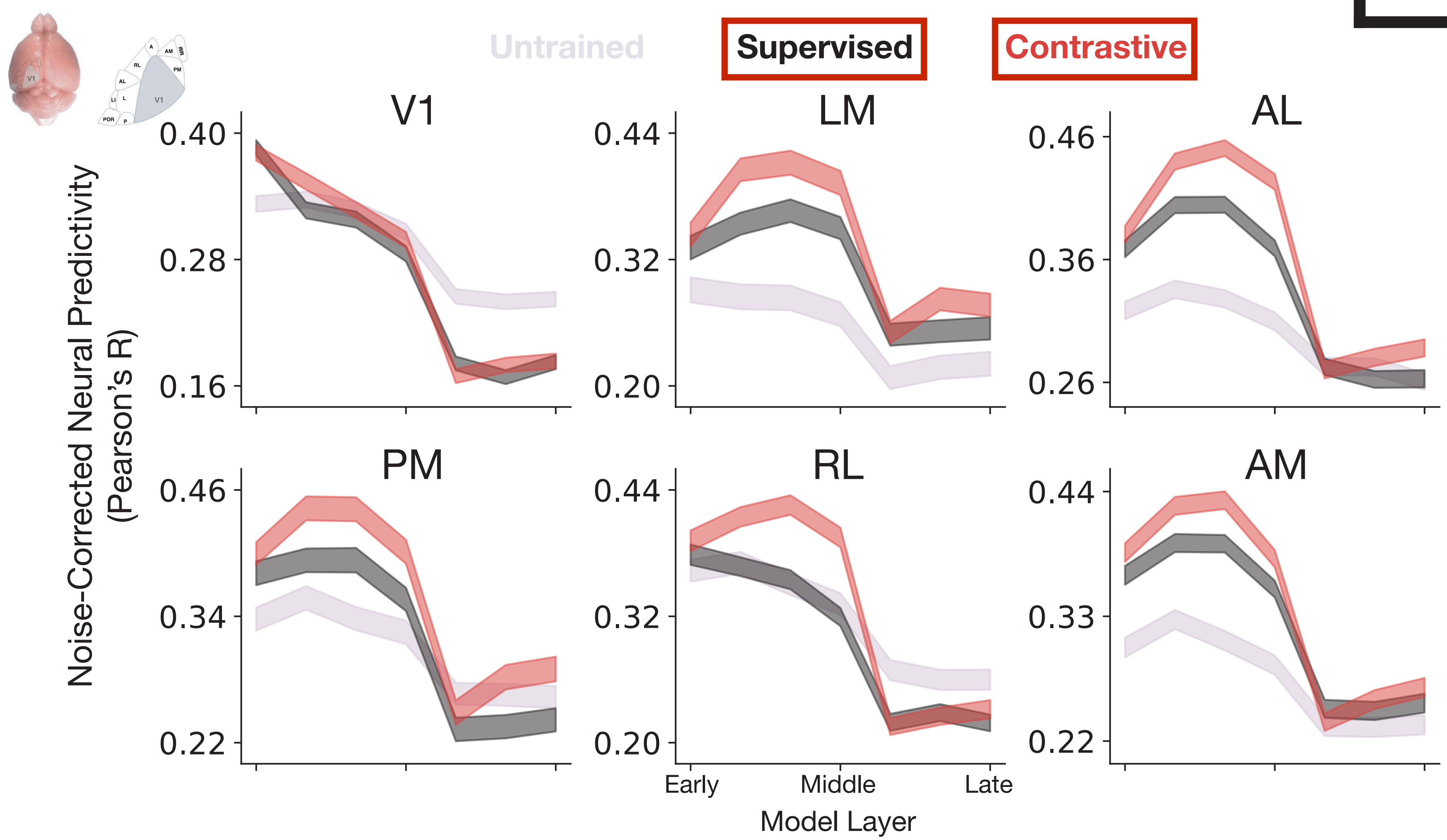
Next Steps: Improving Perceptual Intelligence

A. Nayebi*, N. Kong* *et al.*
Mouse visual cortex as a limited resource system that self-learns an ecologically-general representation. *bioRxiv (2021)*



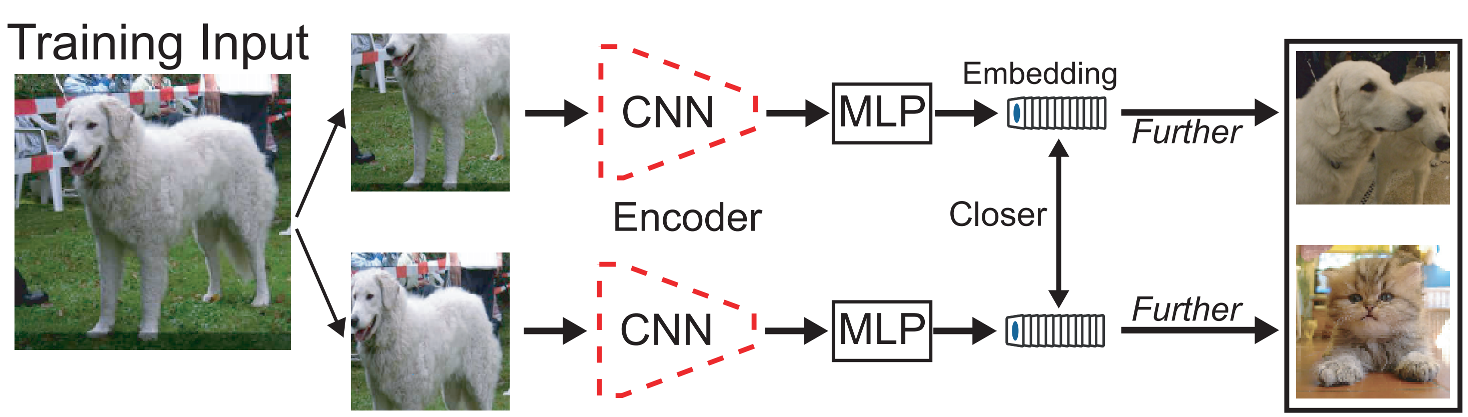
CNN: Convolutional Neural Network, MLP: Multi-Layer Perceptron



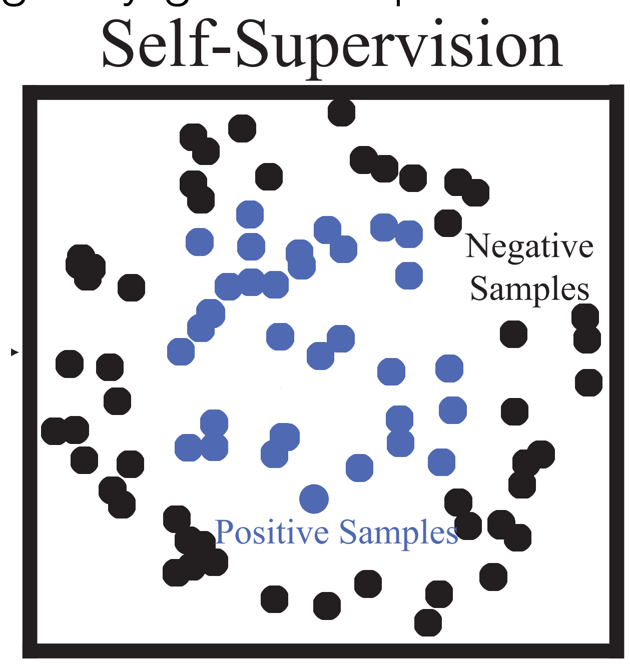


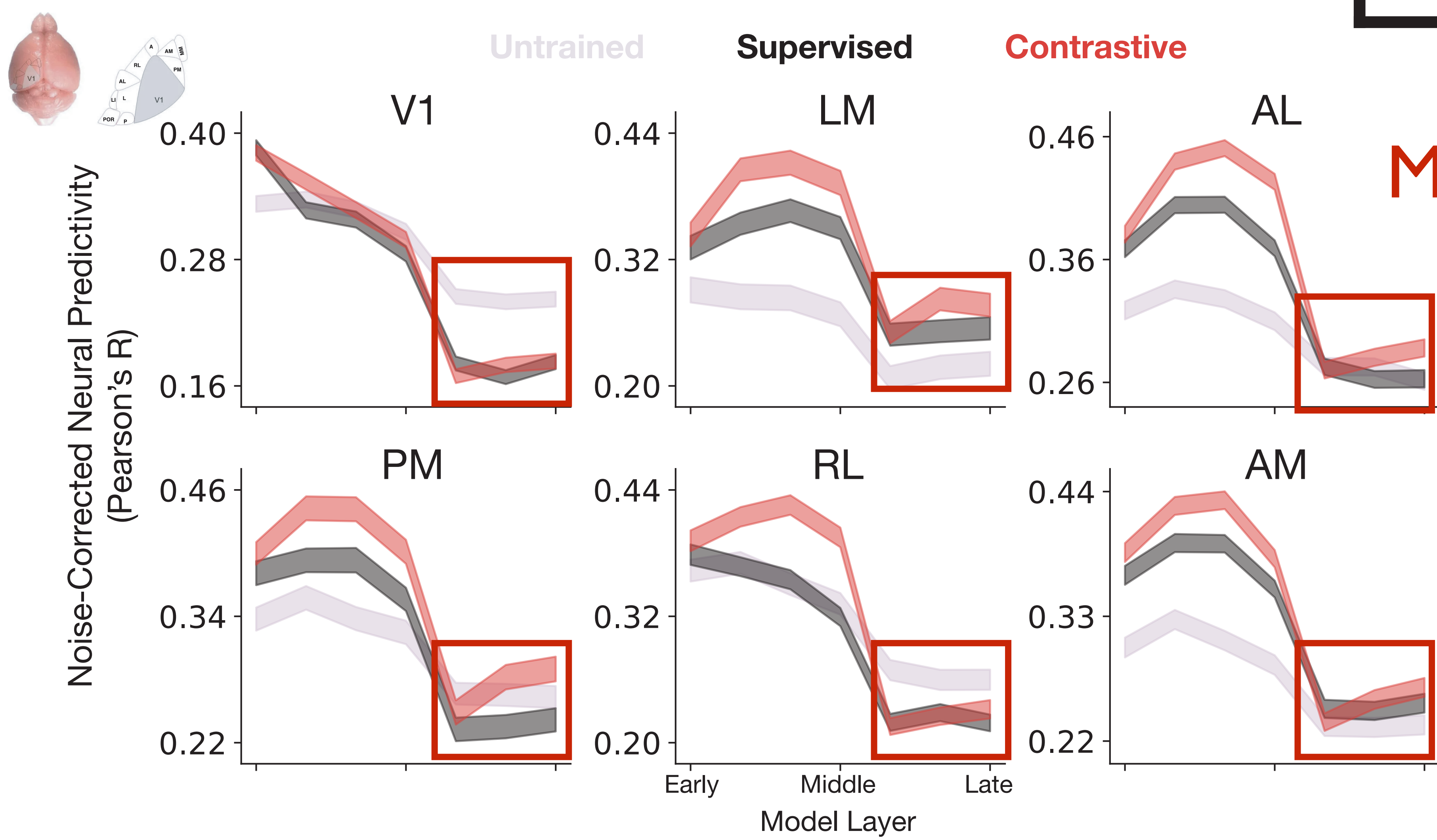
Next Steps: Improving Perceptual Intelligence

A. Nayebi*, N. Kong* *et al.*
Mouse visual cortex as a limited resource system that self-learns an ecologically-general representation. *bioRxiv (2021)*



CNN: Convolutional Neural Network, MLP: Multi-Layer Perceptron

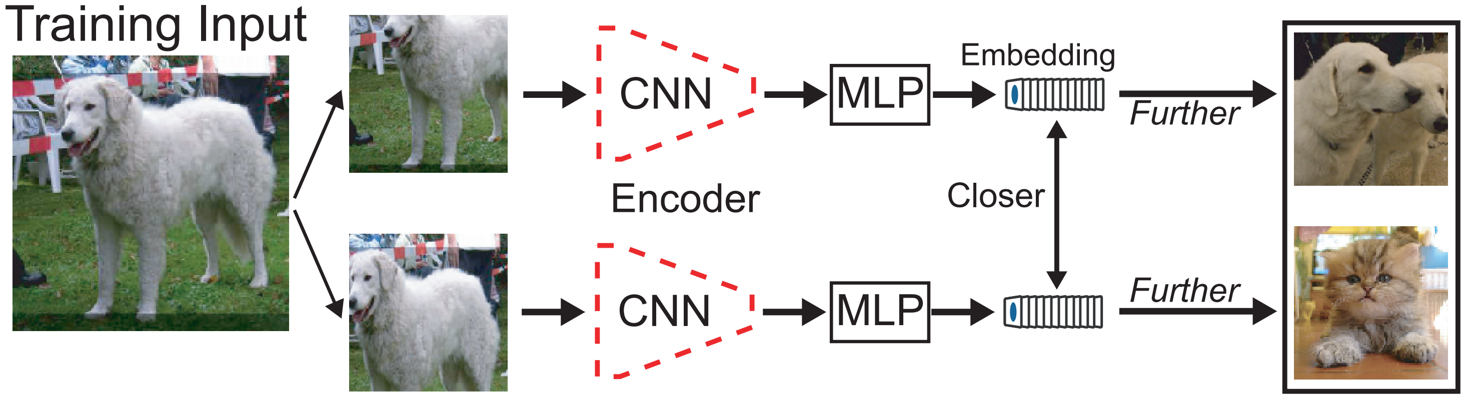




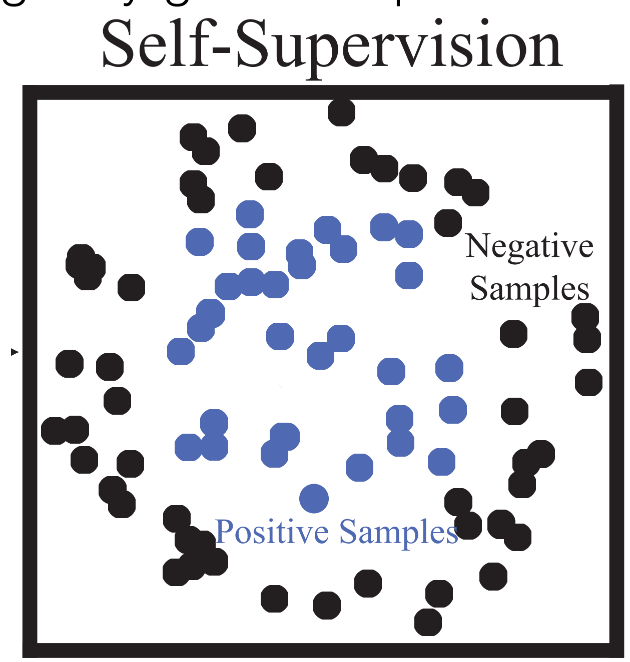
Models are still too deep

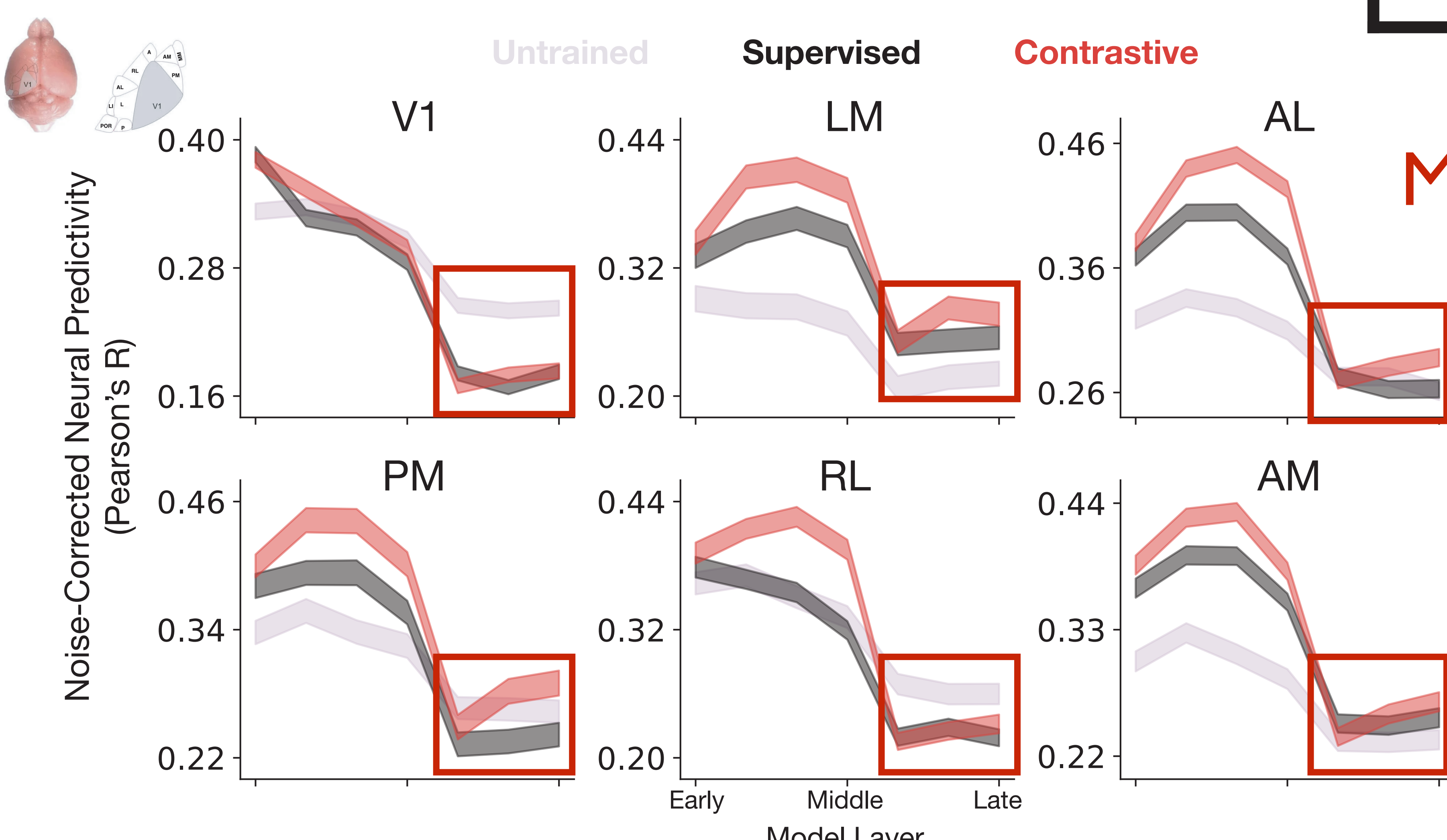
Next Steps: Improving Perceptual Intelligence

A. Nayebi*, N. Kong* *et al.*
Mouse visual cortex as a limited resource system that self-learns an ecologically-general representation. *bioRxiv (2021)*



CNN: Convolutional Neural Network, MLP: Multi-Layer Perceptron



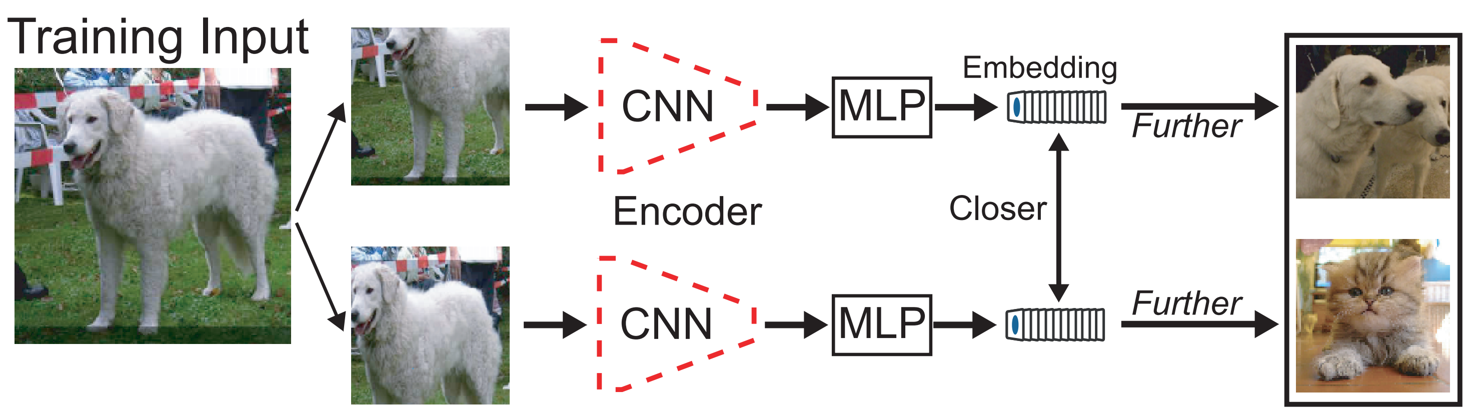


Models are still too deep

Shallower, feedforward models aren't sufficient either!

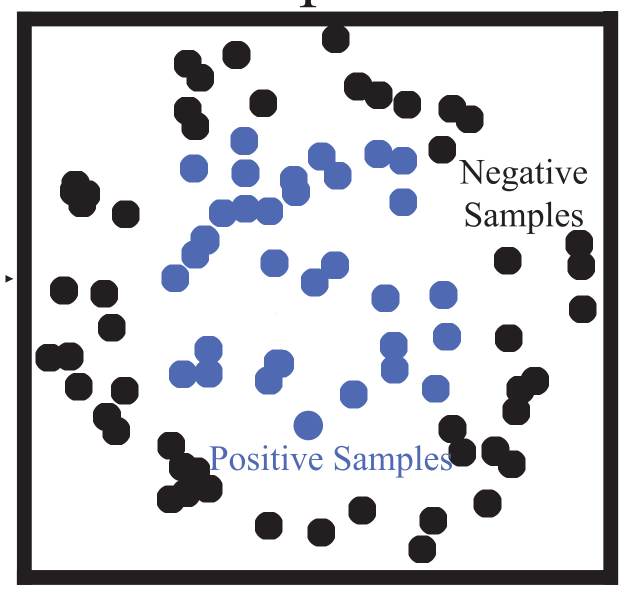
Next Steps: Improving Perceptual Intelligence

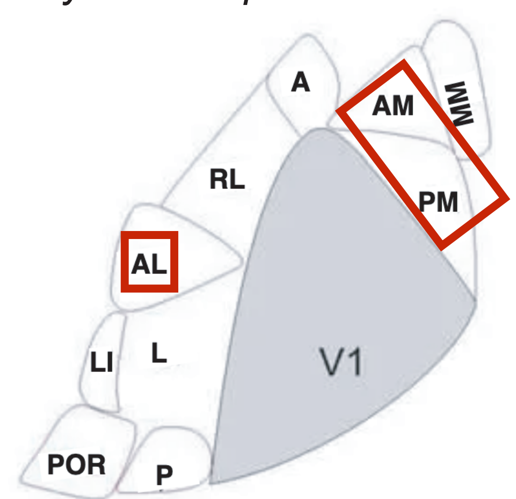
A. Nayebi*, N. Kong* *et al.*
Mouse visual cortex as a limited resource system that self-learns an ecologically-general representation. *bioRxiv (2021)*



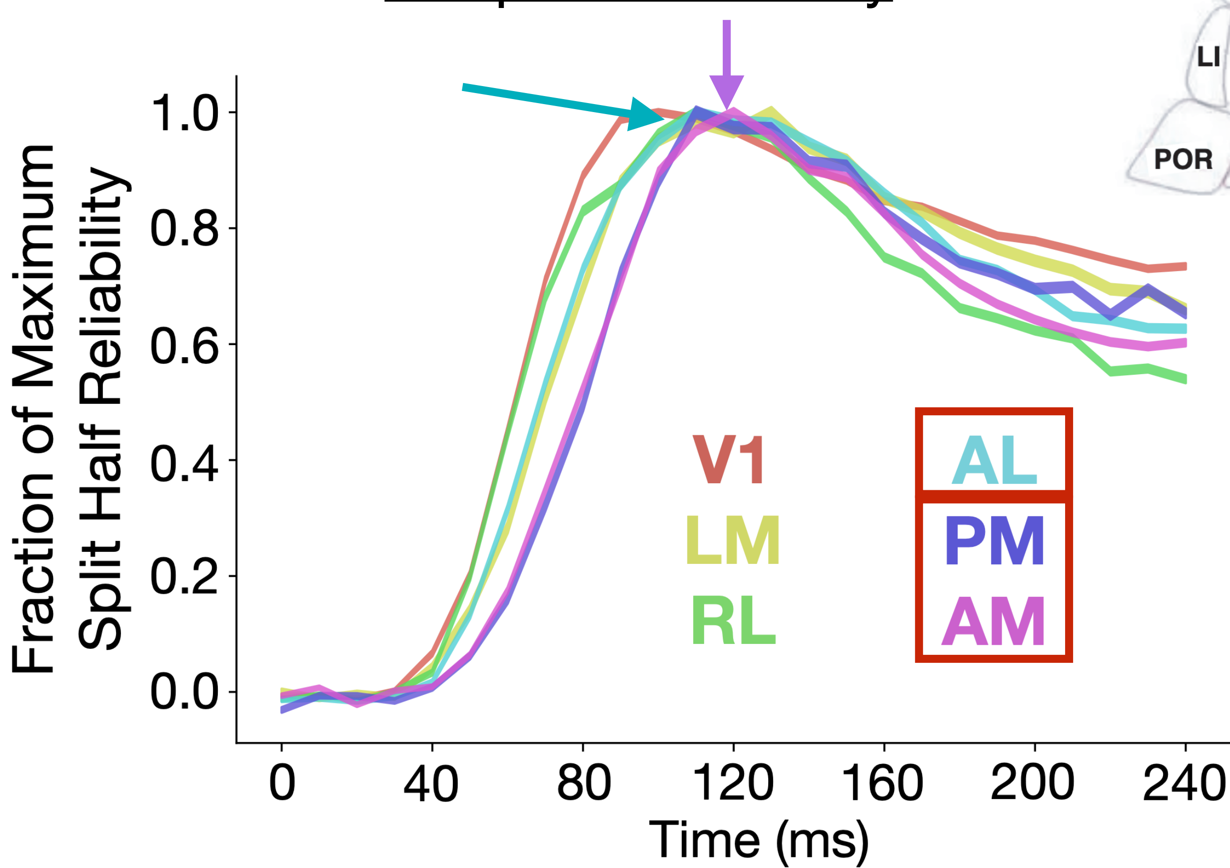
CNN: Convolutional Neural Network, MLP: Multi-Layer Perceptron

Self-Supervision

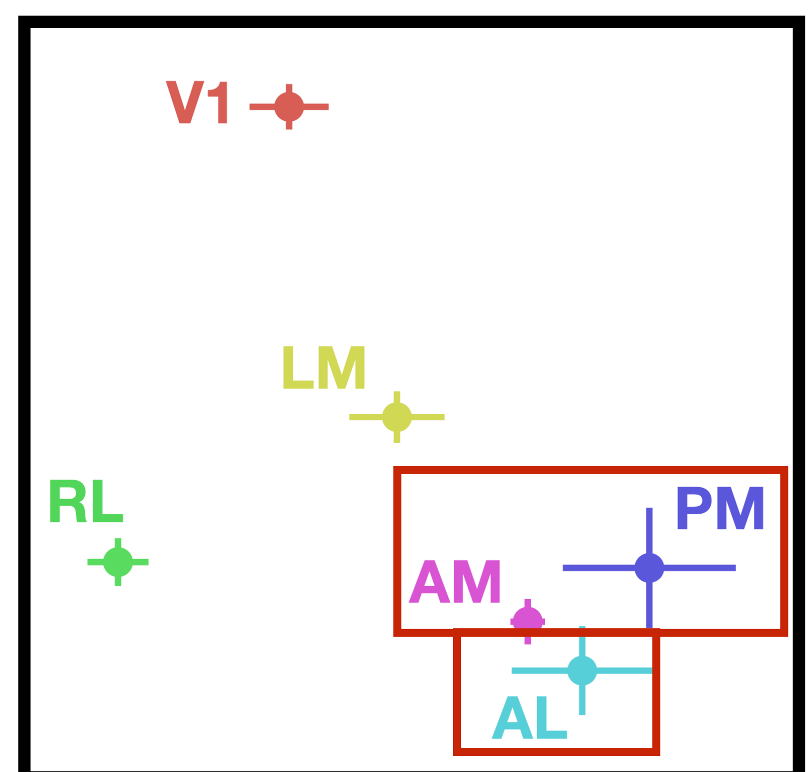




Temporal Hierarchy



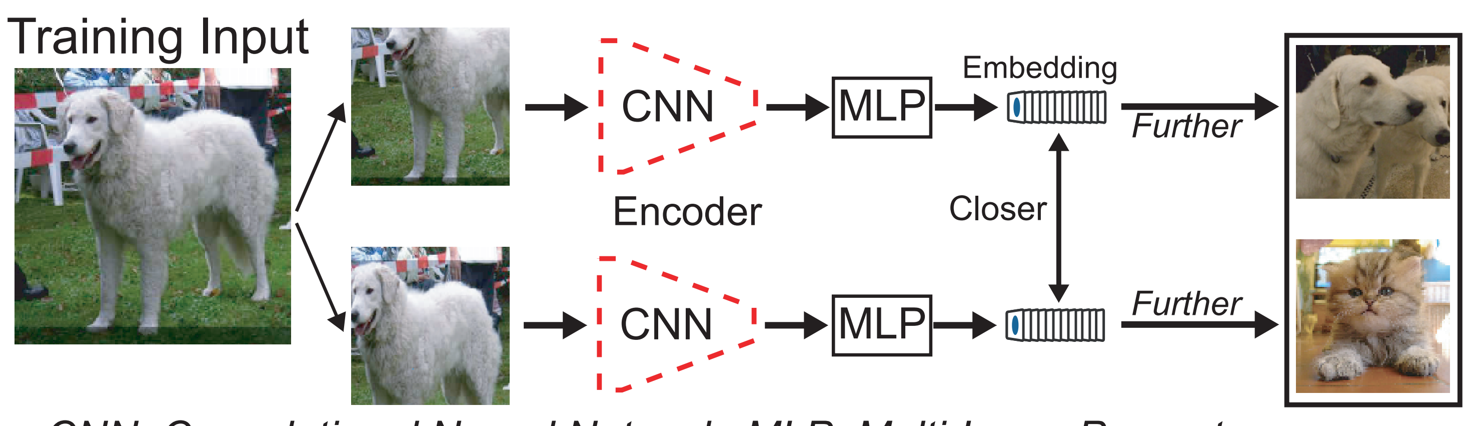
Functional Hierarchy
2D MDS Projection



Next Steps: Improving Perceptual Intelligence

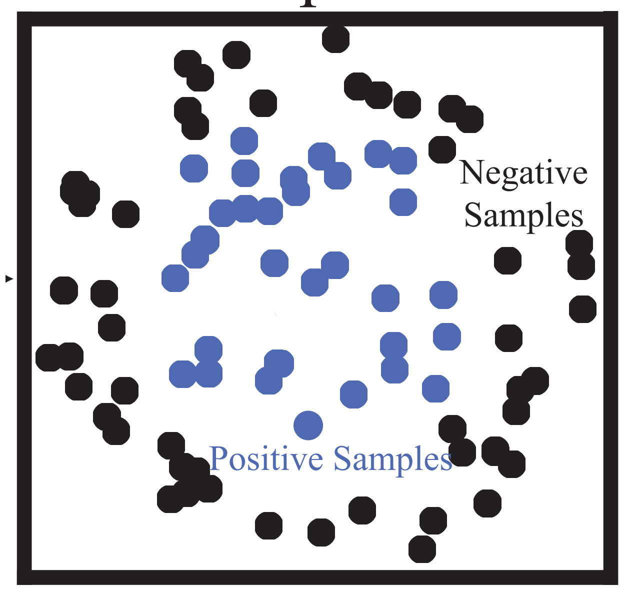
A. Nayebi*, N. Kong* *et al.*

Mouse visual cortex as a limited resource system that self-learns an ecologically-general representation. *bioRxiv (2021)*



CNN: Convolutional Neural Network, MLP: Multi-Layer Perceptron

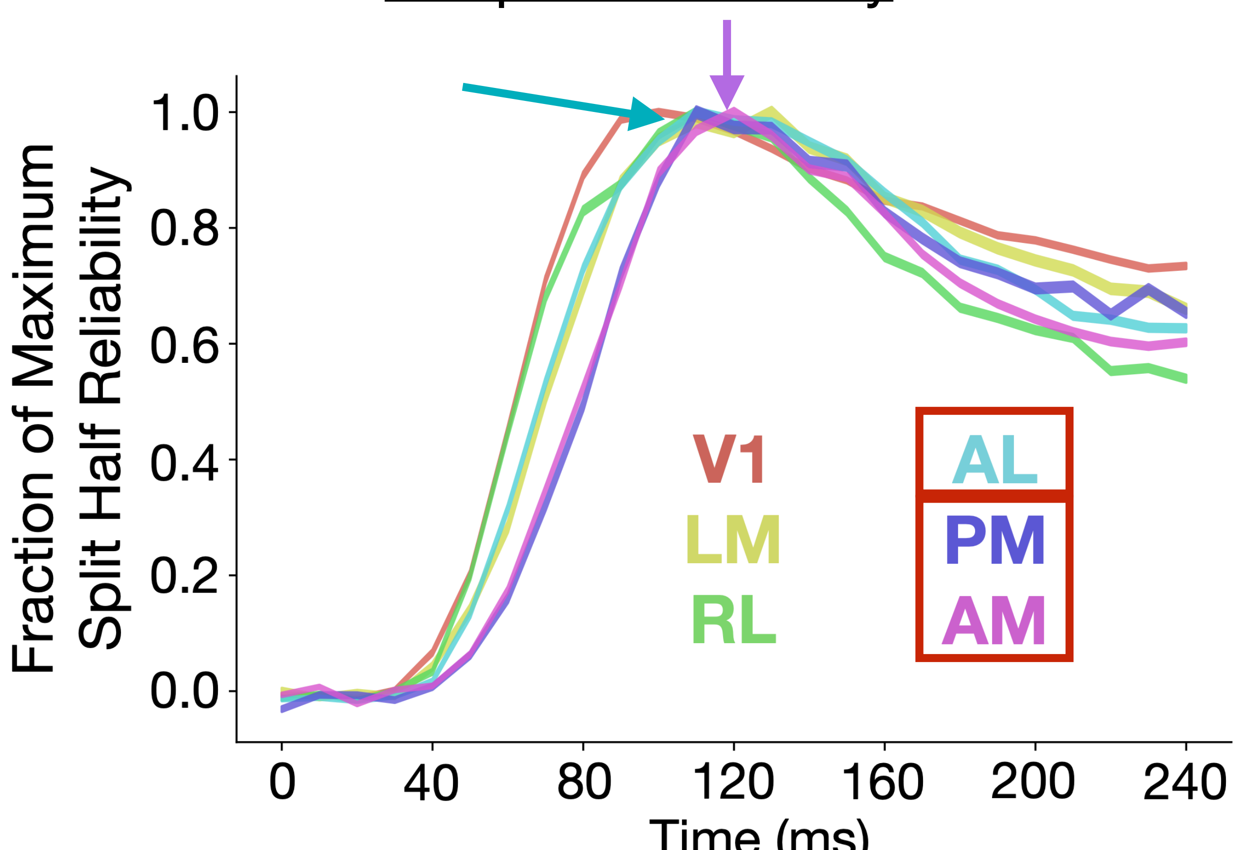
Self-Supervision



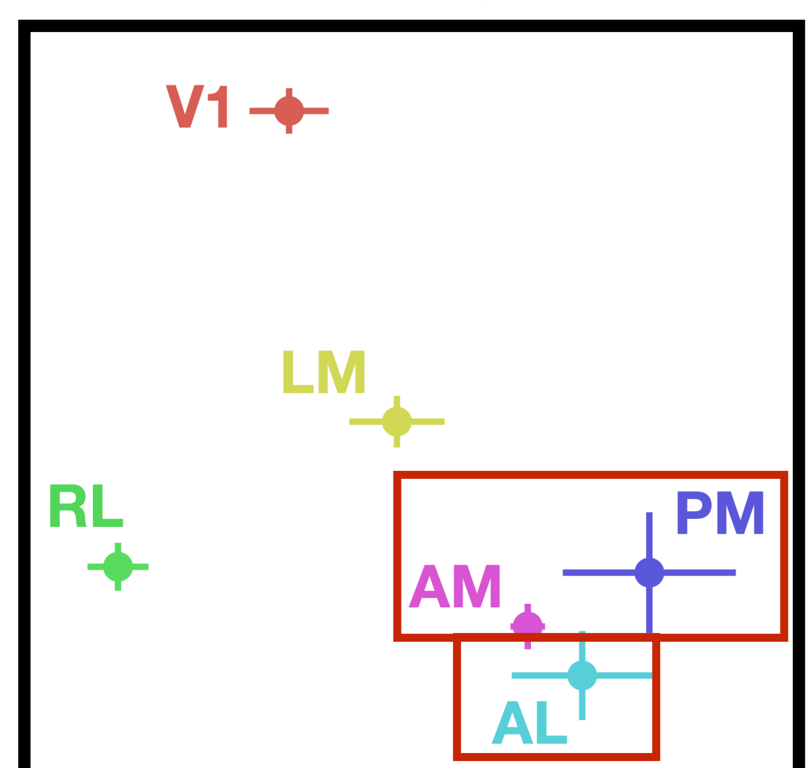
Core Unsolved Question:

Circuit busting with a shallower, *recurrent* network?

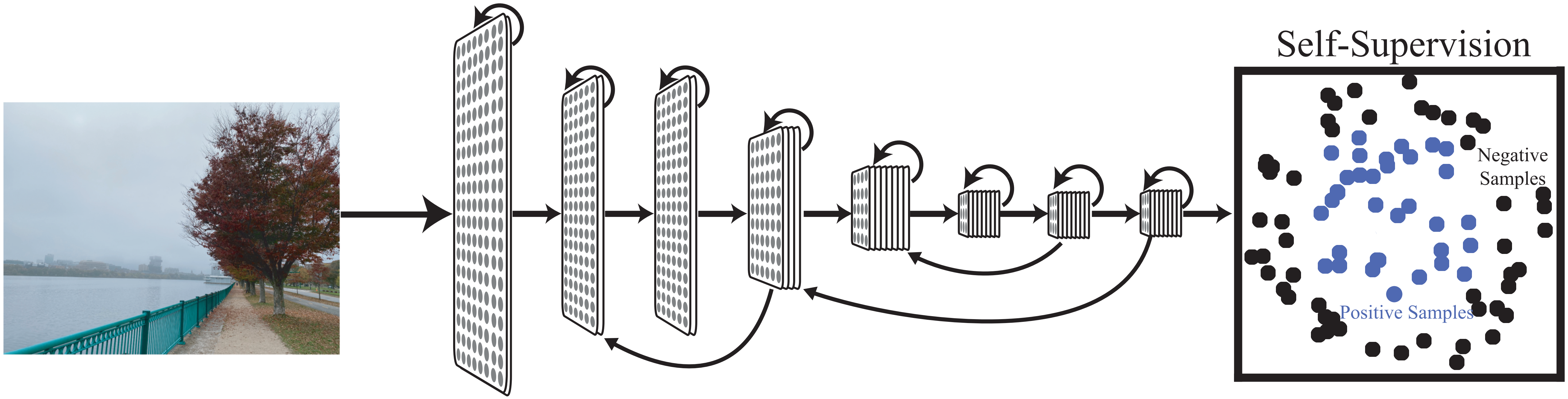
Temporal Hierarchy



Functional Hierarchy
2D MDS Projection

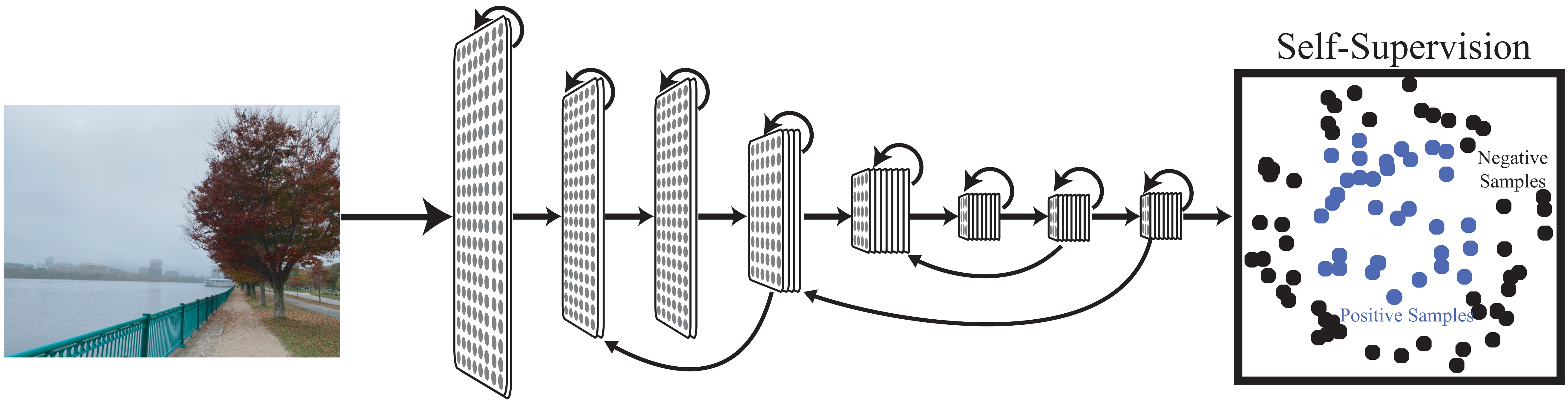


Combined Recurrent, Self-Supervised models



Combined recurrent, self-supervised models

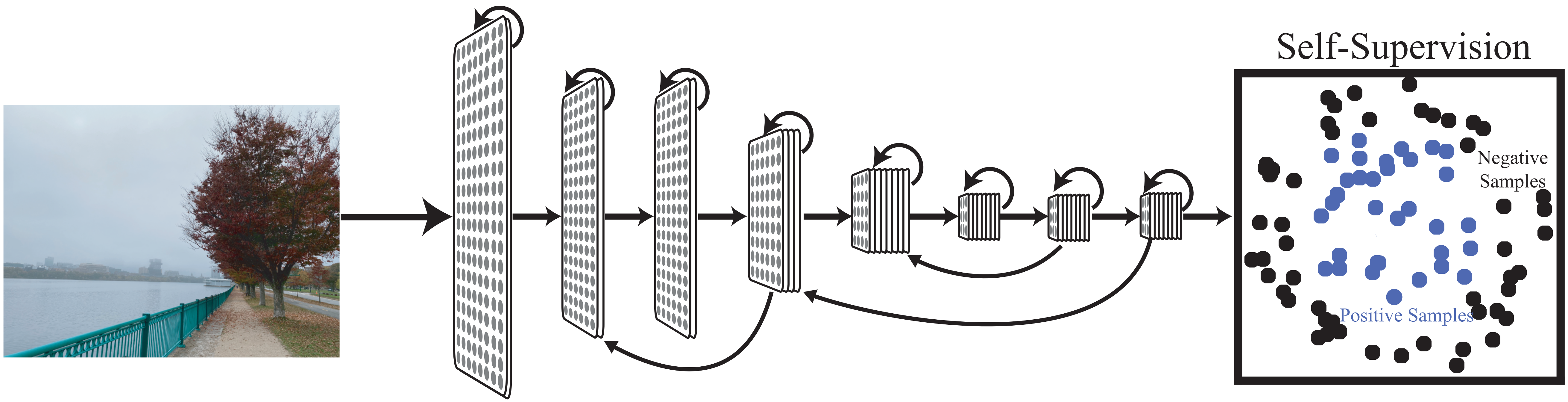
Combined Recurrent, Self-Supervised models



Combined recurrent, self-supervised models

Directly interfaces with the dynamic inputs animals naturally encounter

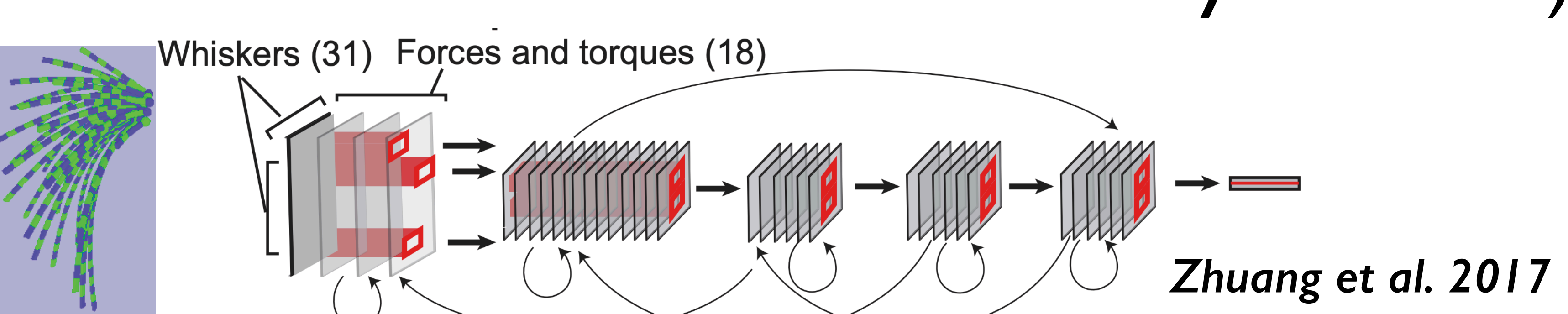
Combined Recurrent, Self-Supervised models



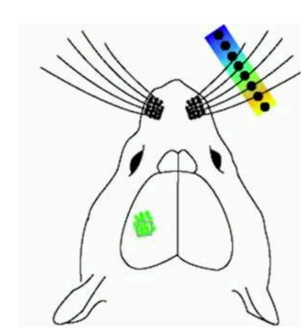
Combined recurrent, self-supervised models

Directly interfaces with the dynamic inputs animals naturally encounter

Parsimony across sensory systems? (e.g. auditory, barrel, and olfactory cortices)

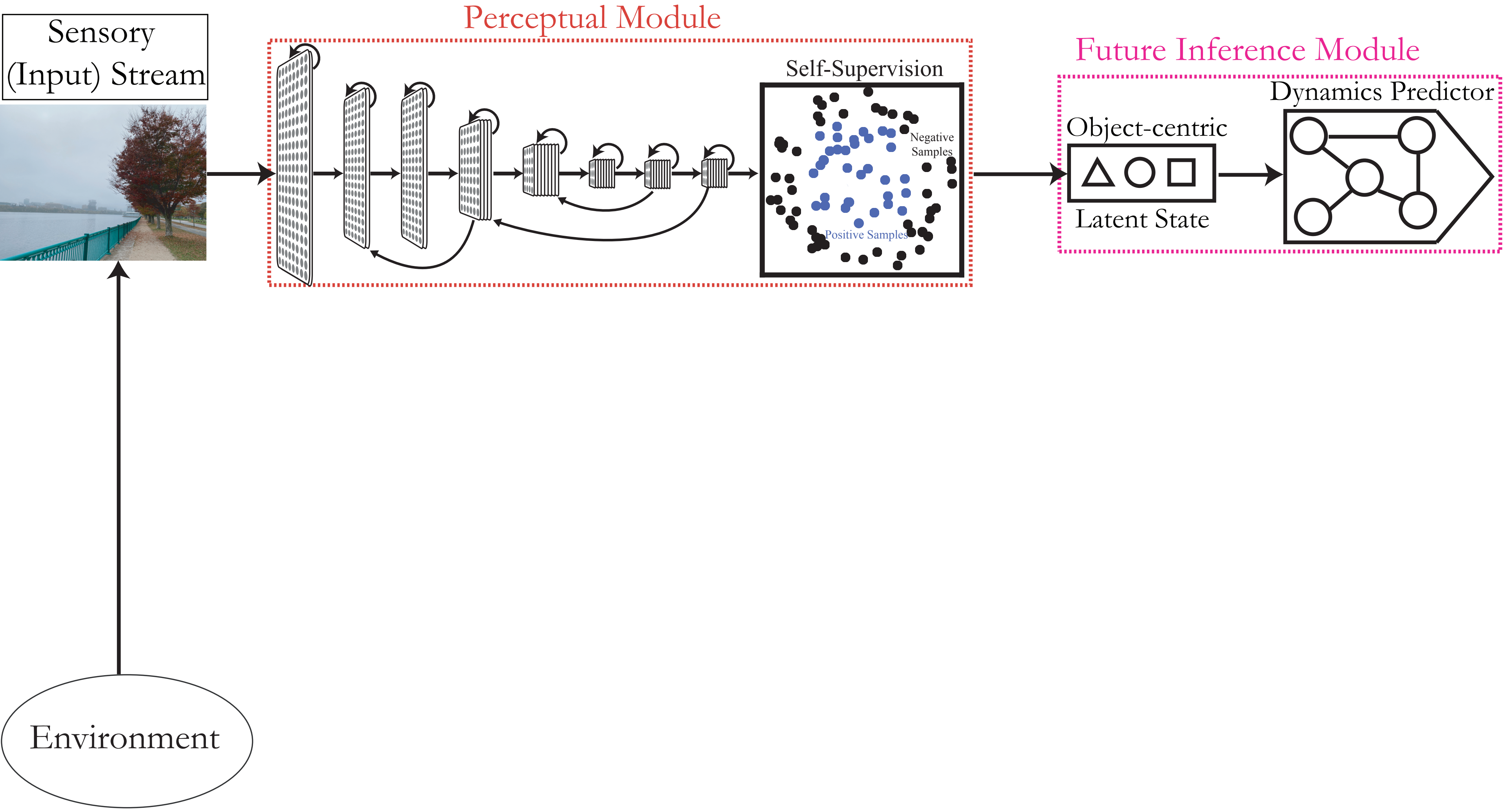


Zhuang et al. 2017



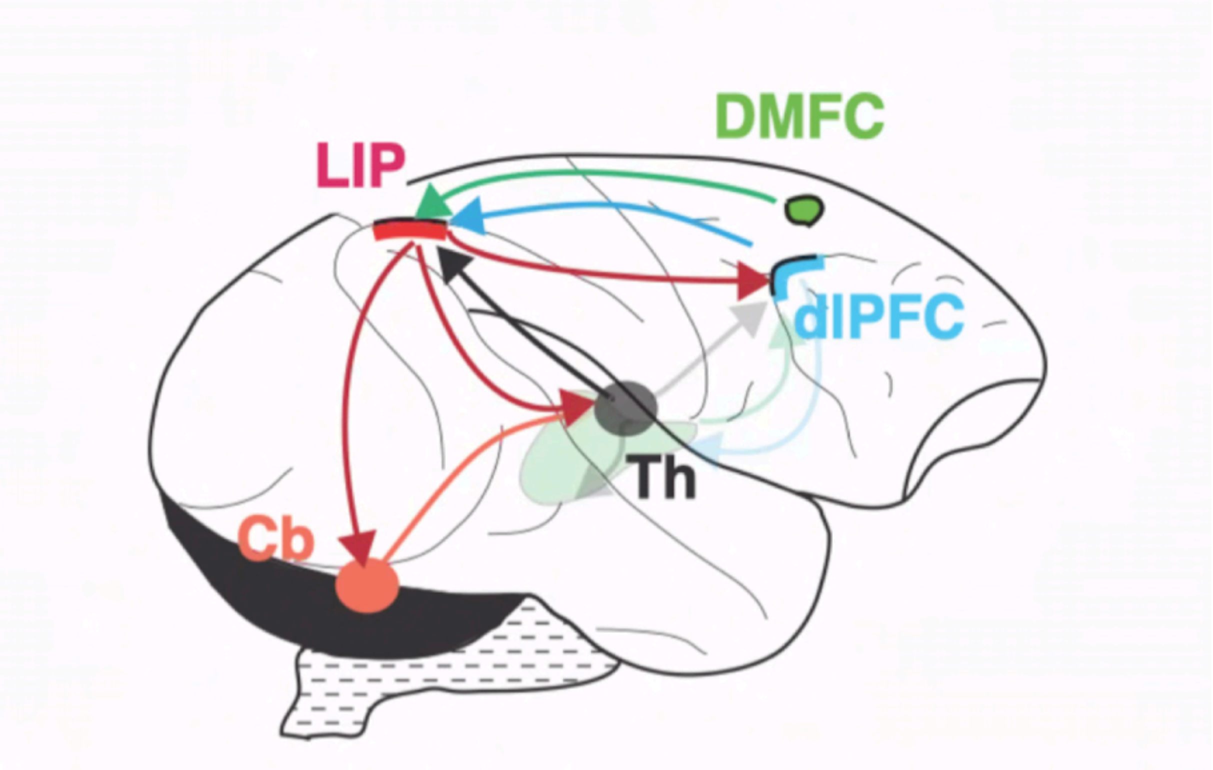
Feather et al. 2020

Next Steps: Computational Principles of Future Inference



Next Steps: Computational Principles of Future Inference

What mechanisms enable the FPN to generally perform mental simulation?

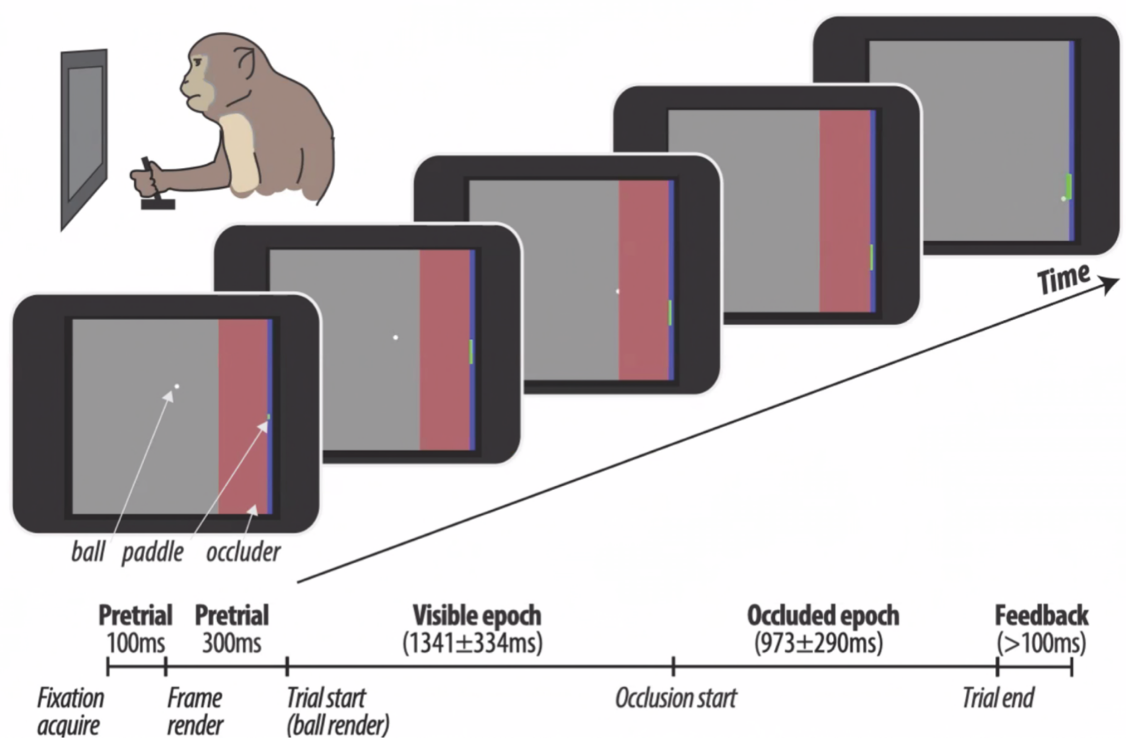


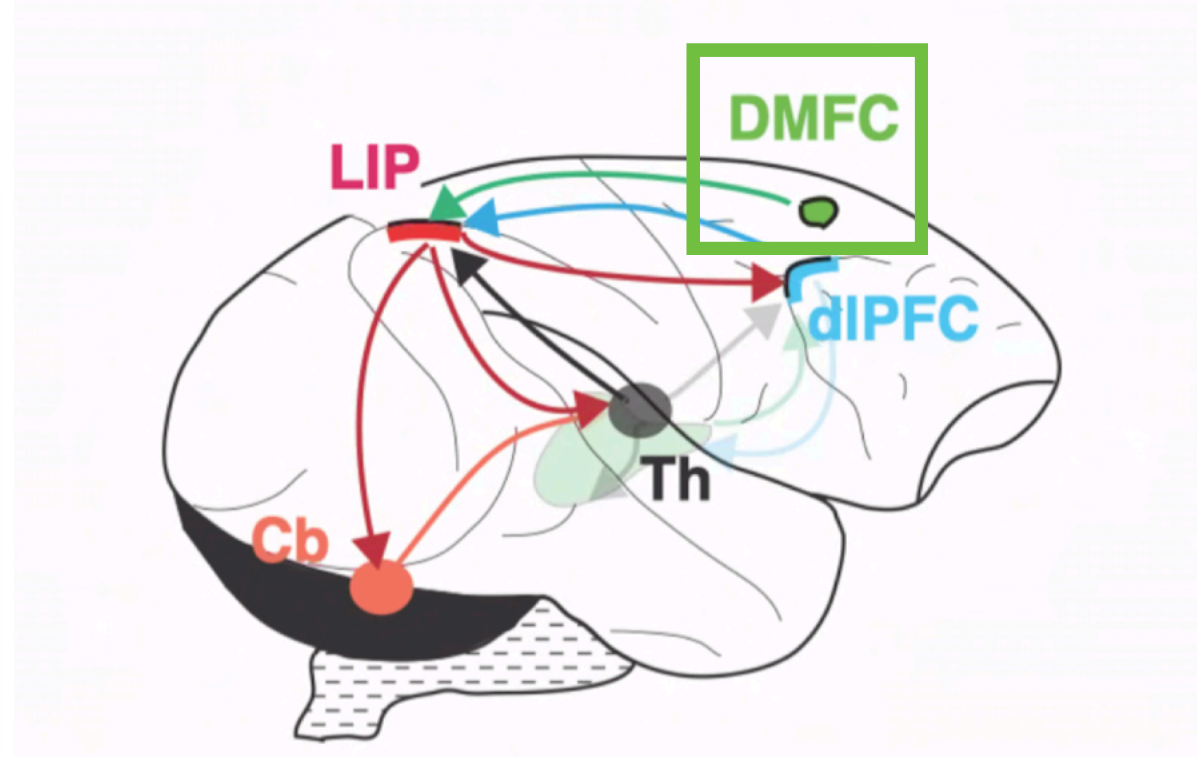
Fronto-Parietal Network (FPN)

Next Steps: Computational Principles of Future Inference

What mechanisms enable the FPN to generally perform mental simulation?

Behavioral task (Mental-Pong)





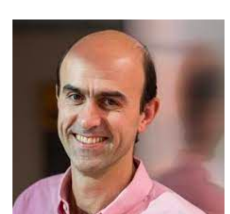
Fronto-Parietal Network (FPN)

The role of mental simulation in primate physical inference abilities

Rishi Rajalingham, Aida Piccato, Mehrdad Jazayeri

doi: https://doi.org/10.1101/2021.01.14.426741

Rishi Rajalingham



Mehrdad Jazayeri

Dynamic tracking of objects in the macaque dorsomedial frontal cortex

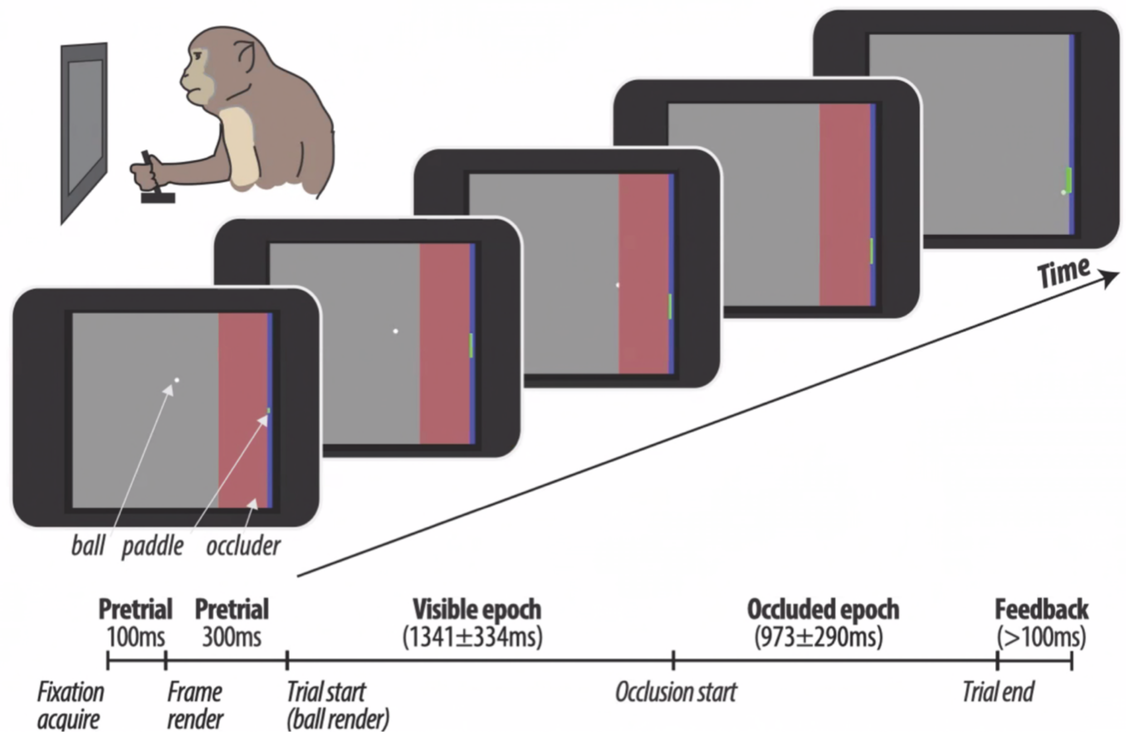
Rishi Rajalingham, Hansem Sohn, Mehrdad Jazayeri

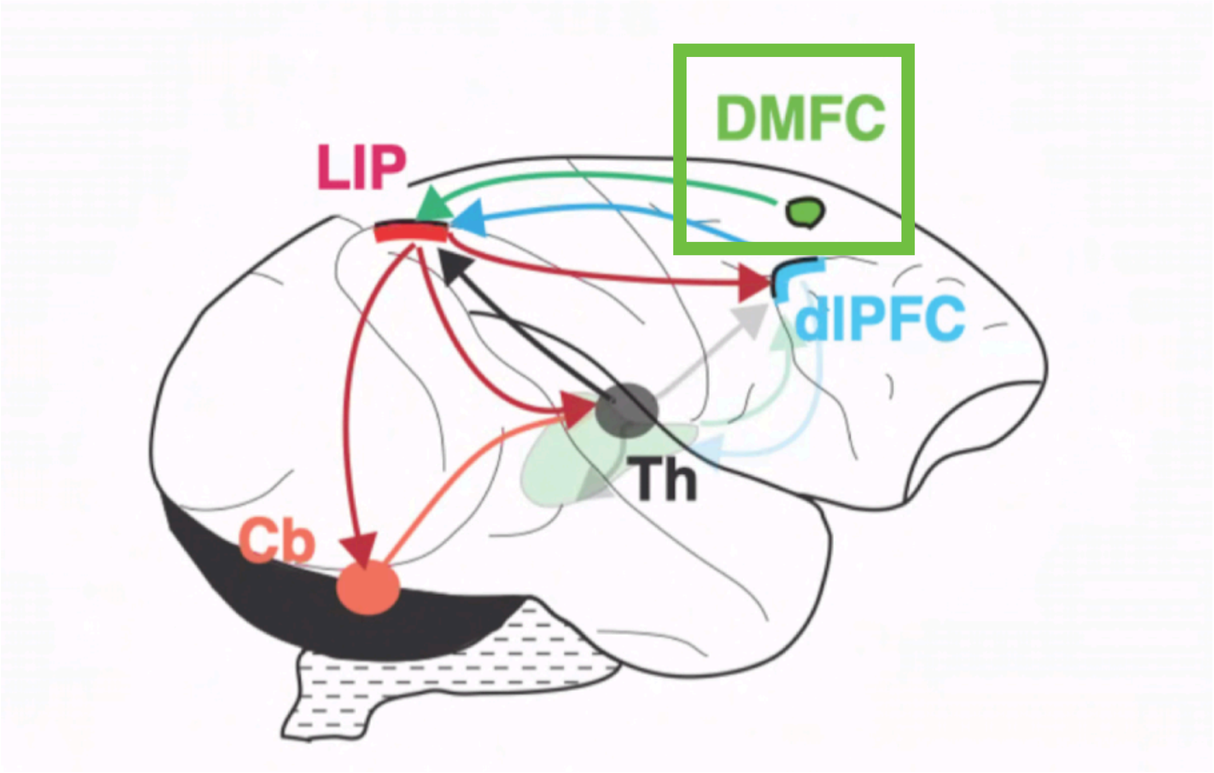
doi: https://doi.org/10.1101/2022.06.24.497529

Next Steps: Computational Principles of Future Inference

What mechanisms enable the FPN to generally perform mental simulation?

Behavioral task (Mental-Pong)





Fronto-Parietal Network (FPN)

Many algorithmic choices to explore: Are mental simulations object-centric, scene-centric, or more fine-grained?

The role of mental simulation in primate physical inference abilities

Rishi Rajalingham, Aida Piccato, Mehrdad Jazayeri

doi: https://doi.org/10.1101/2021.01.14.426741


Rishi Rajalingham


Mehrdad Jazayeri

Dynamic tracking of objects in the macaque dorsomedial frontal cortex

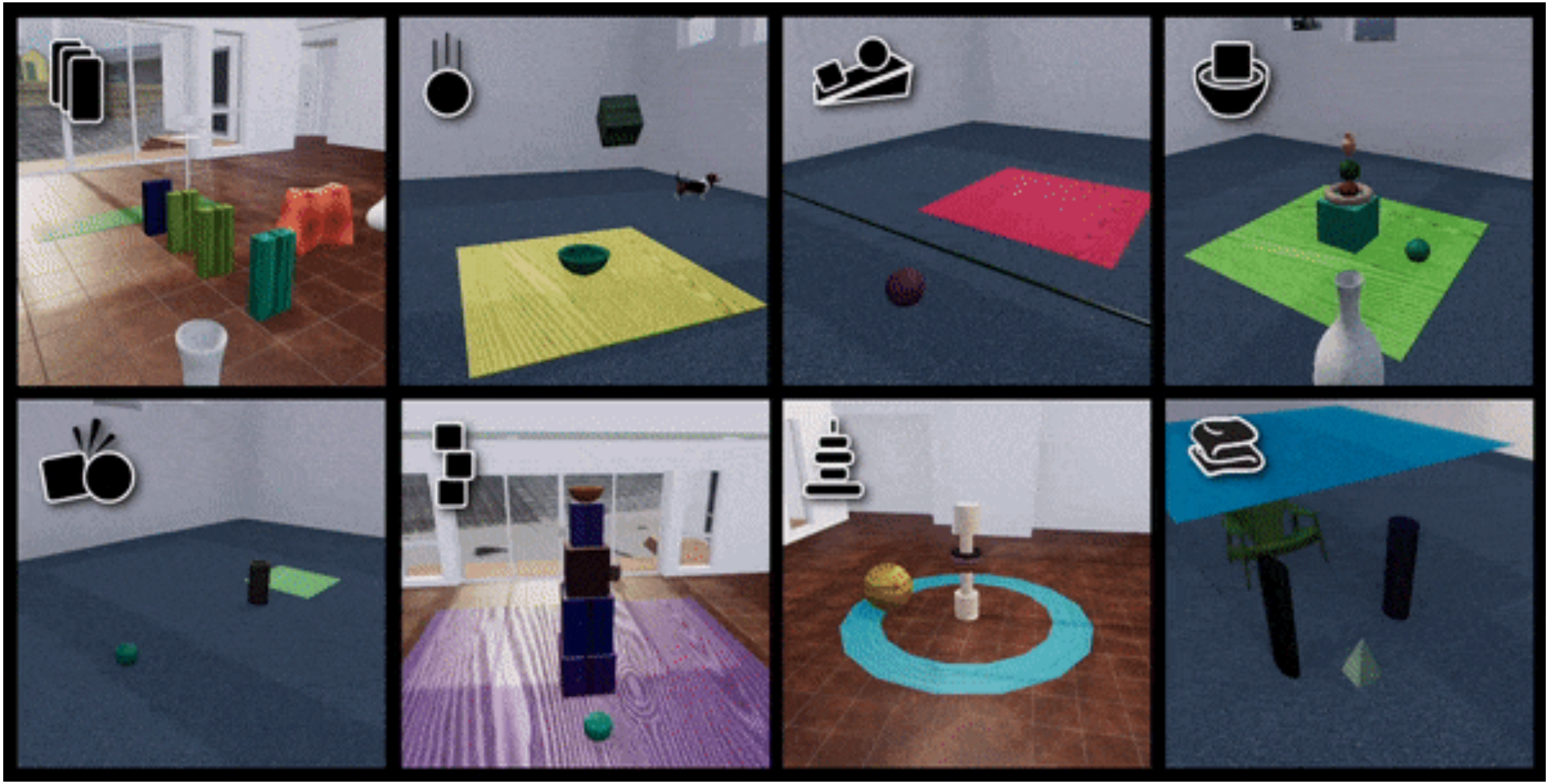
Rishi Rajalingham, Hansem Sohn, Mehrdad Jazayeri

doi: https://doi.org/10.1101/2022.06.24.497529

Dynamics Environment

Physion/ThreeDWorld (TDW)

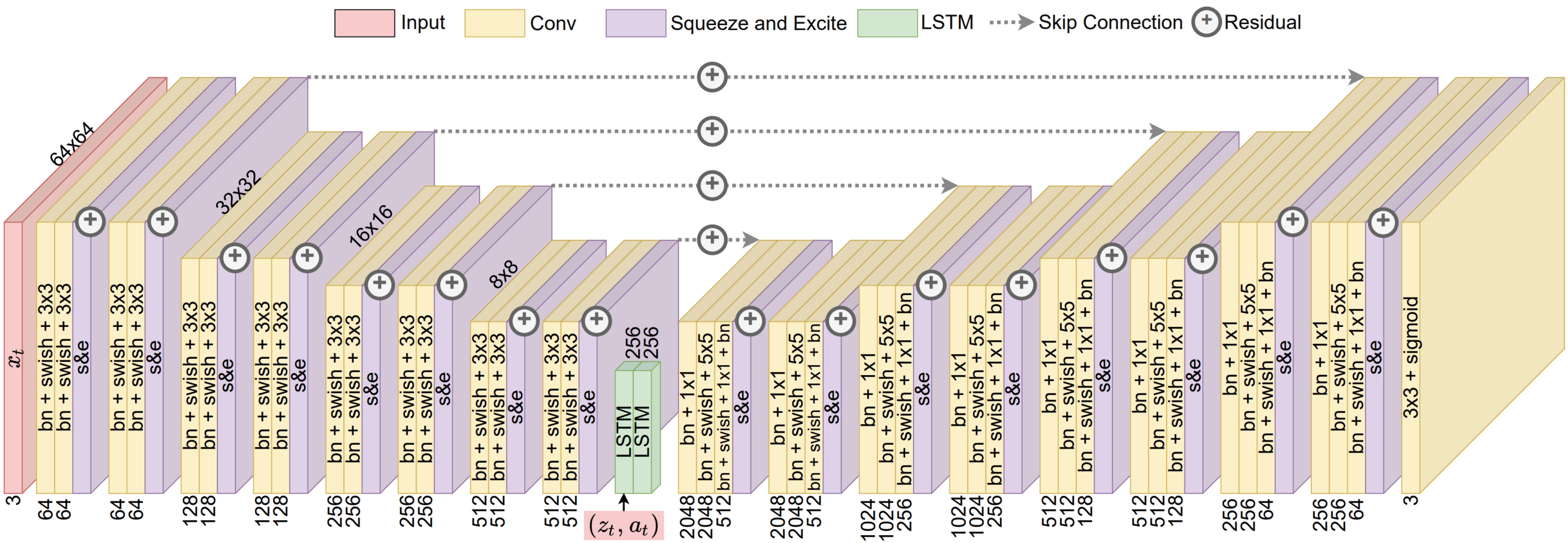
Bear et al. 2021



Focus on everyday physical understanding

Pixel-Wise Frame Prediction: Basic Components

Babaeizadeh et al. 2021

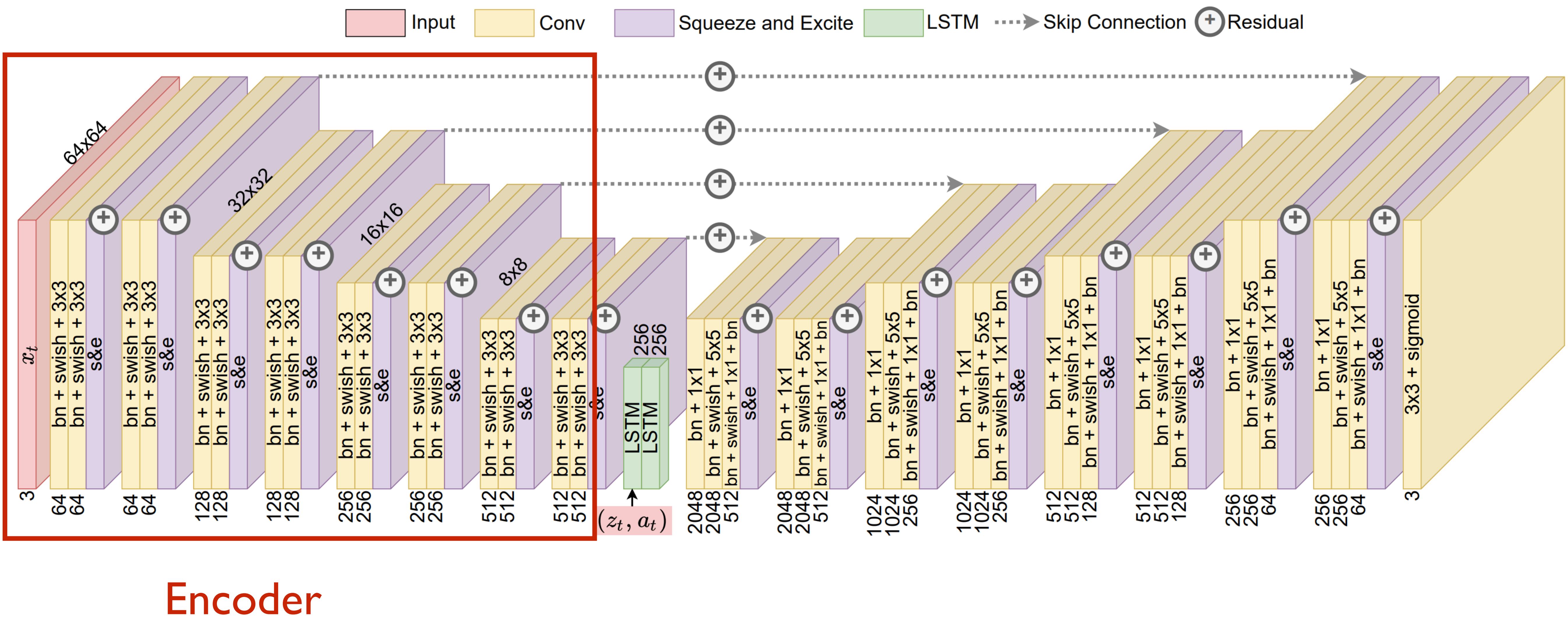


<u>Common AI paradigm:</u>

Self-supervised on future *pixel-wise* frame prediction — can be readily applied to large-scale, real-world video datasets

Pixel-Wise Frame Prediction: Basic Components

Babaeizadeh et al. 2021

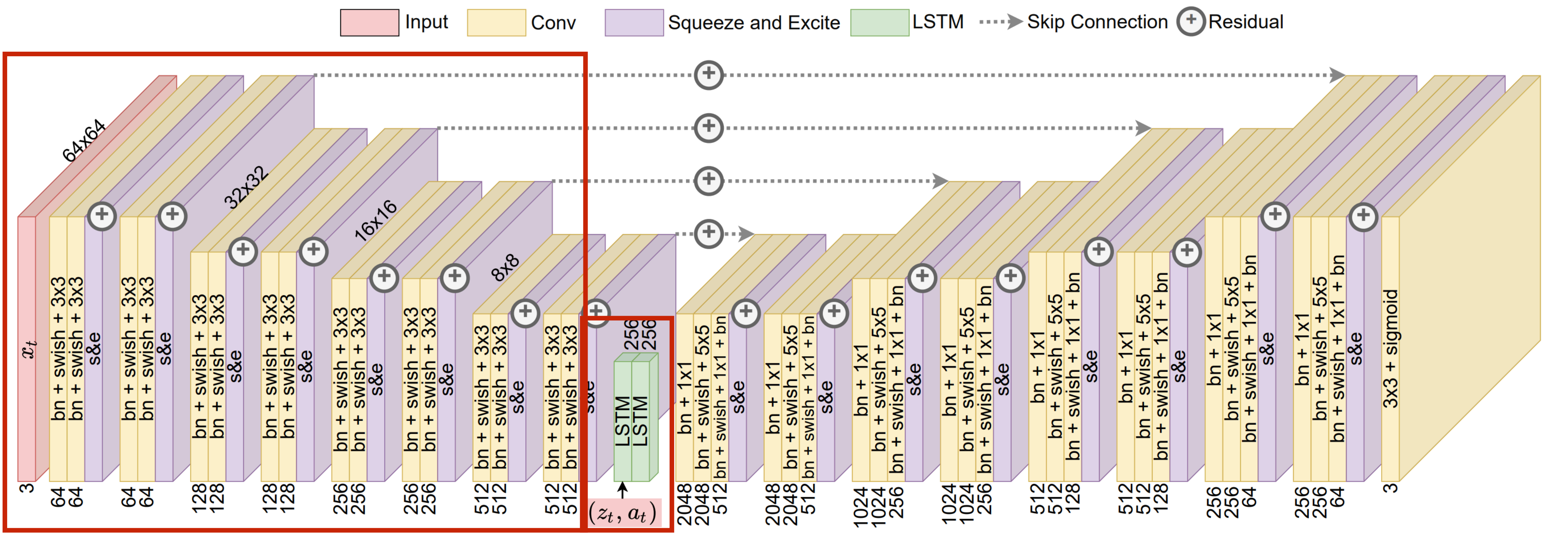


<u>Common AI paradigm:</u>

Self-supervised on future *pixel-wise* frame prediction — can be readily applied to large-scale, real-world video datasets

Pixel-Wise Frame Prediction: Basic Components

Babaeizadeh et al. 2021



Encoder

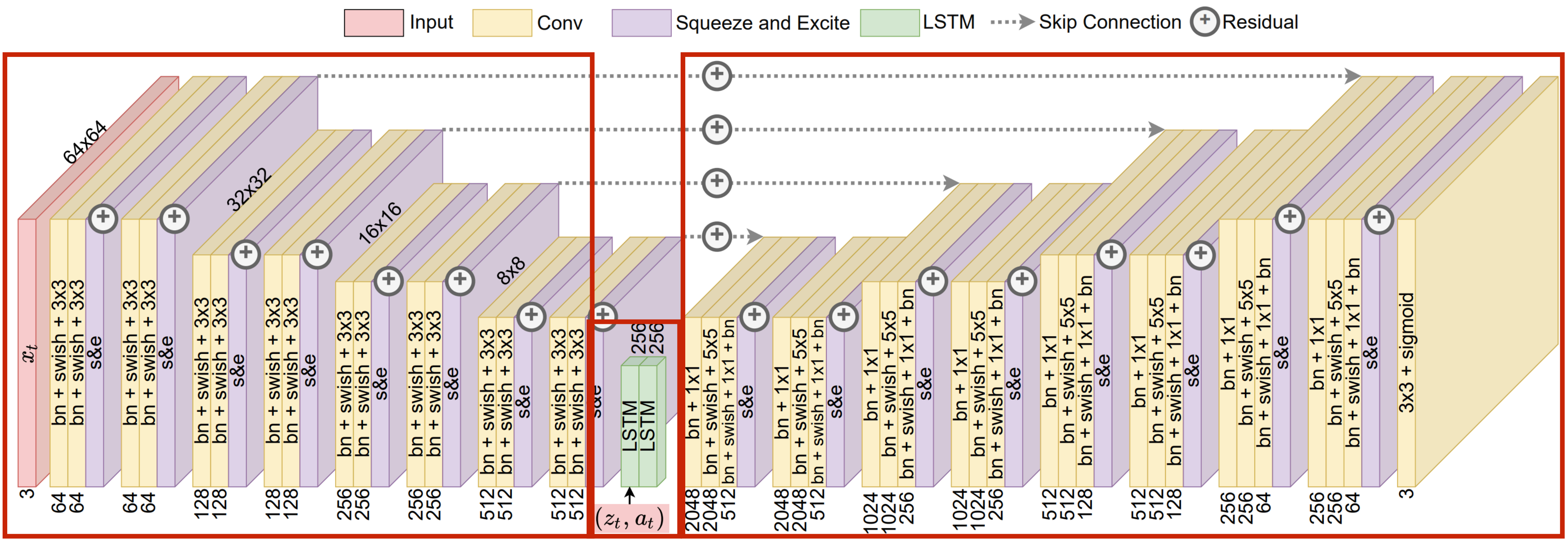
Dynamics Predictor

<u>Common AI paradigm:</u>

Self-supervised on future *pixel-wise* frame prediction — can be readily applied to large-scale, real-world video datasets

Pixel-Wise Frame Prediction: Basic Components

Babaeizadeh et al. 2021



Encoder

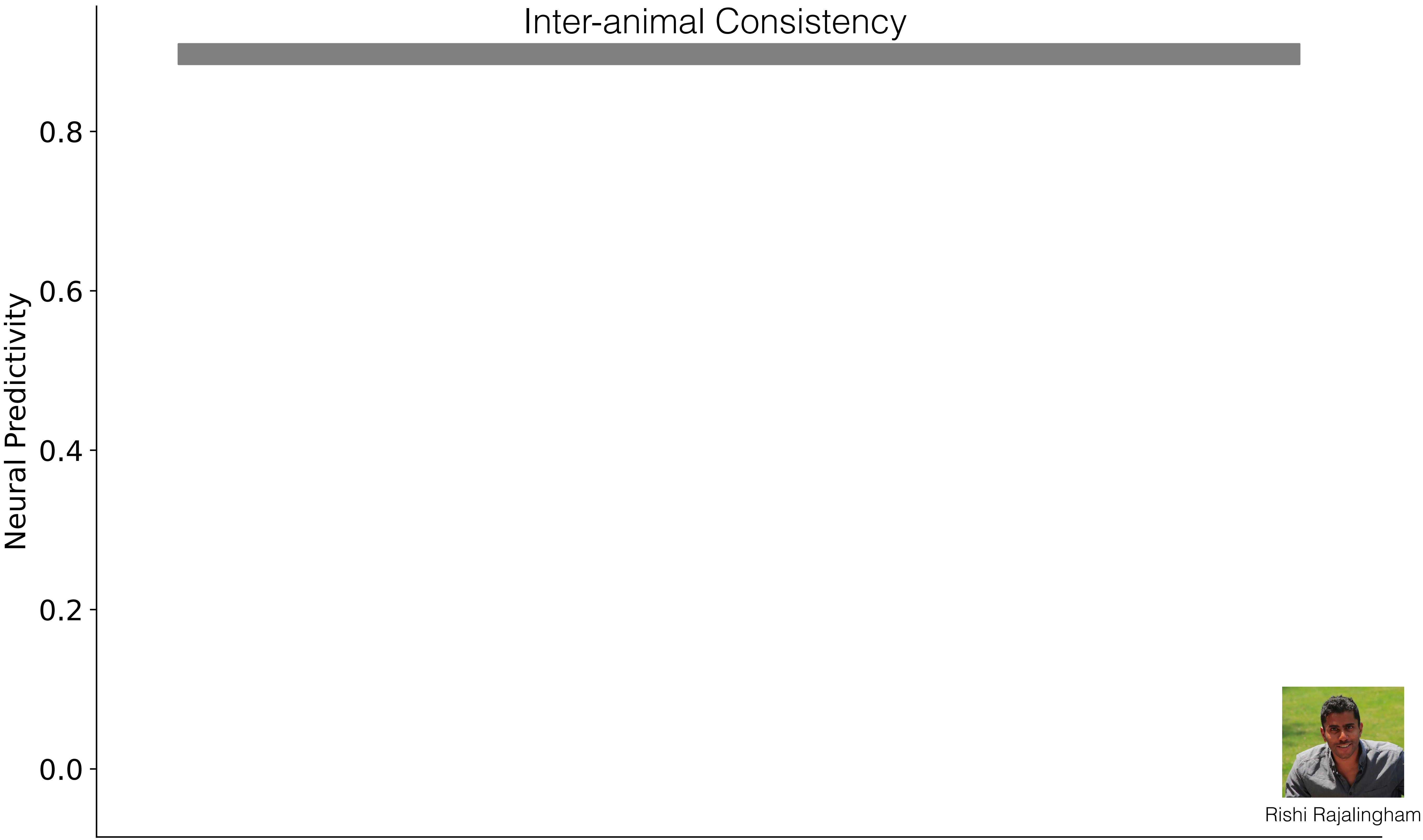
Dynamics Predictor

Decoder (Loss function)

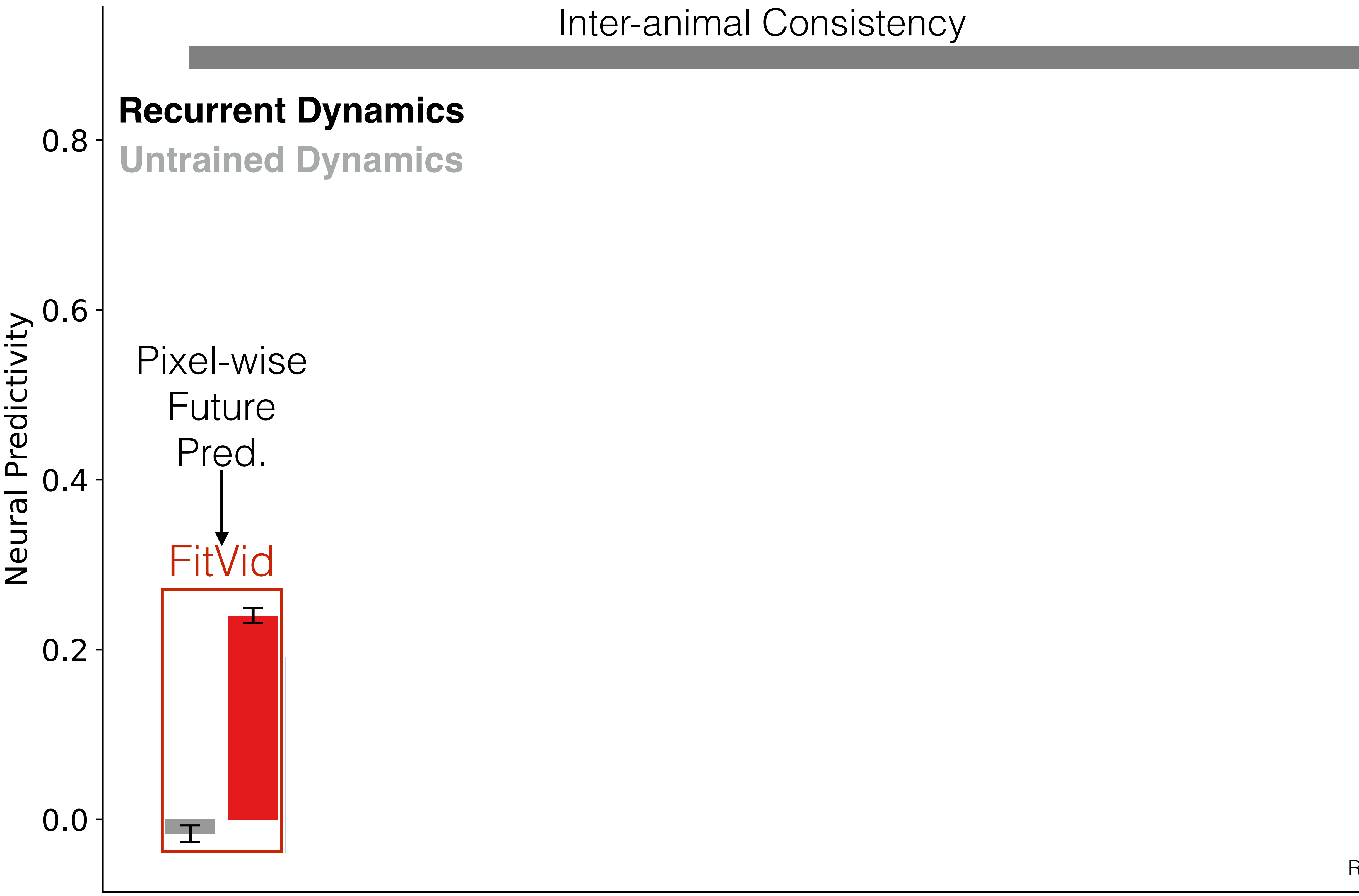
<u>Common AI paradigm:</u>

Self-supervised on future *pixel-wise* frame prediction — can be readily applied to large-scale, real-world video datasets

Preliminary Results Modeling DMFC

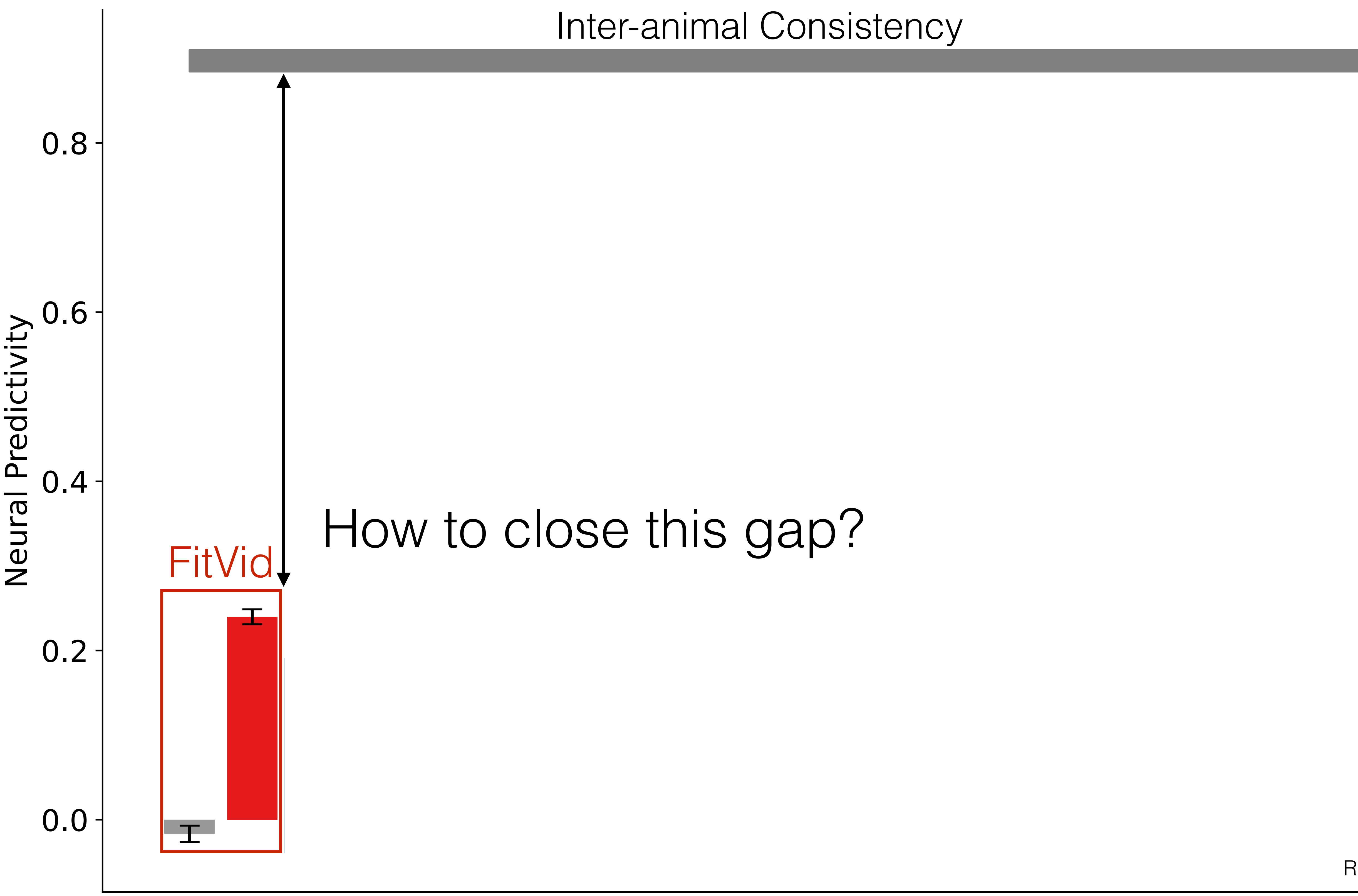


Preliminary Results Modeling DMFC



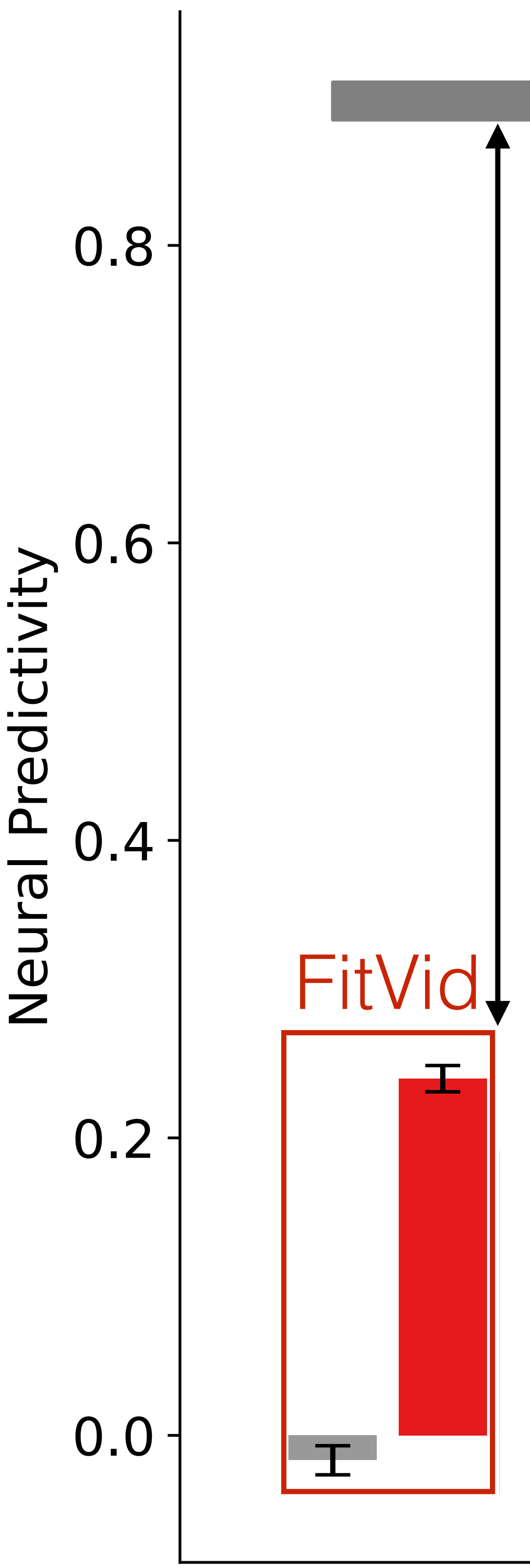
Rishi Rajalingham

Preliminary Results Modeling DMFC



Rishi Rajalingham

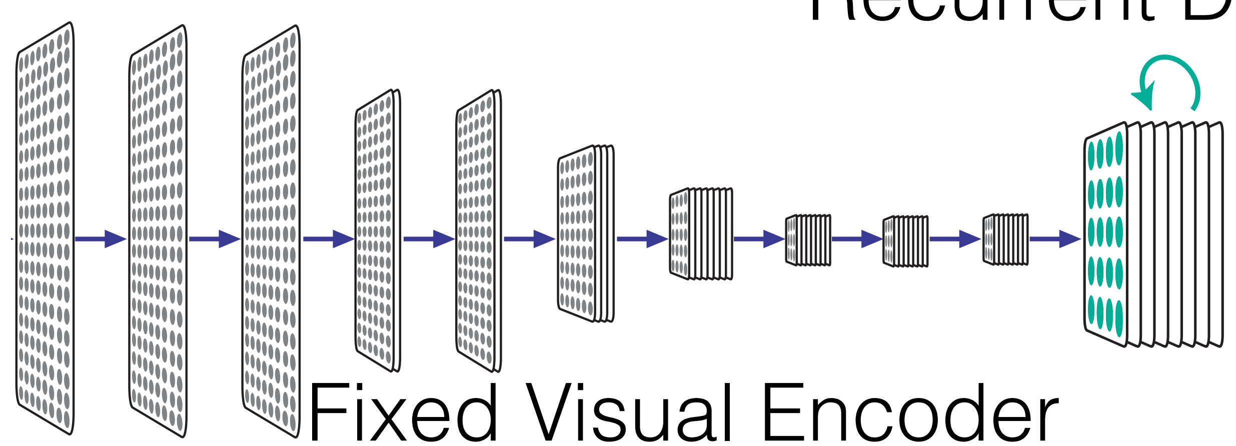
Preliminary Results Modeling DMFC



Alternative Idea:

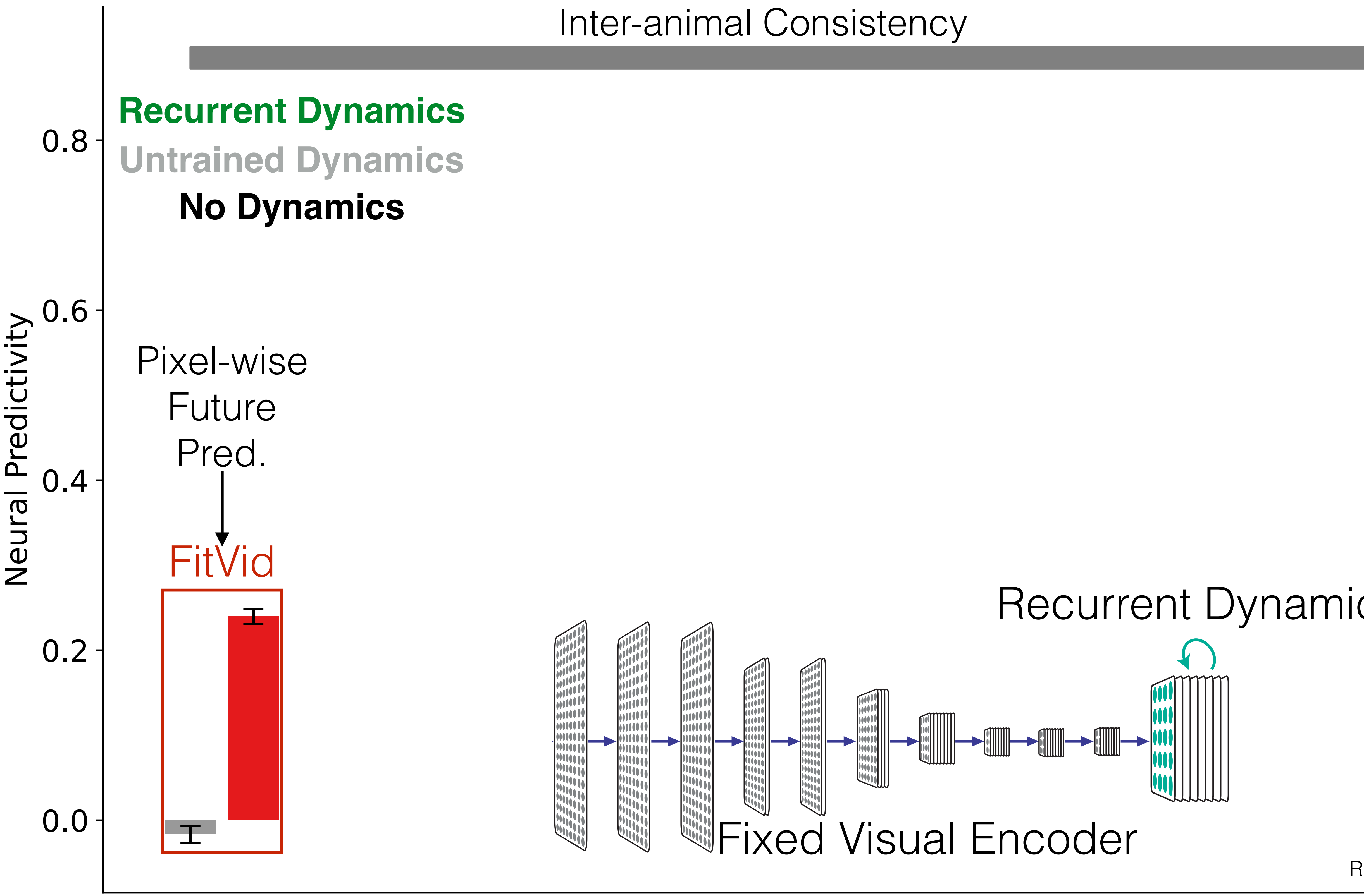
Take a fixed "visual system" pretrained to produce "useful" (for object/scene-understanding) latents, and predict those instead of the pixels

How to close this gap?



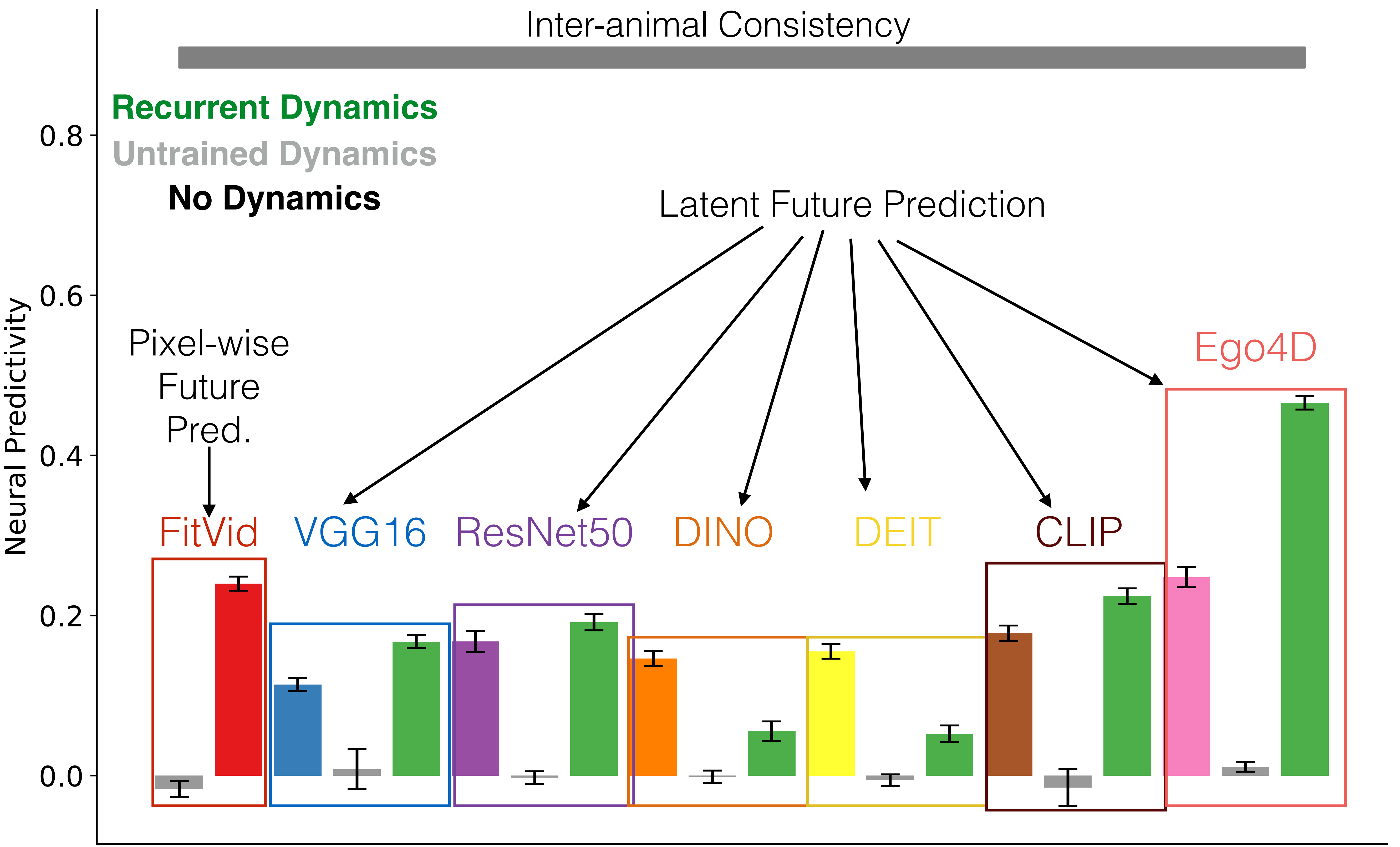
Rishi Rajalingham

Preliminary Results Modeling DMFC

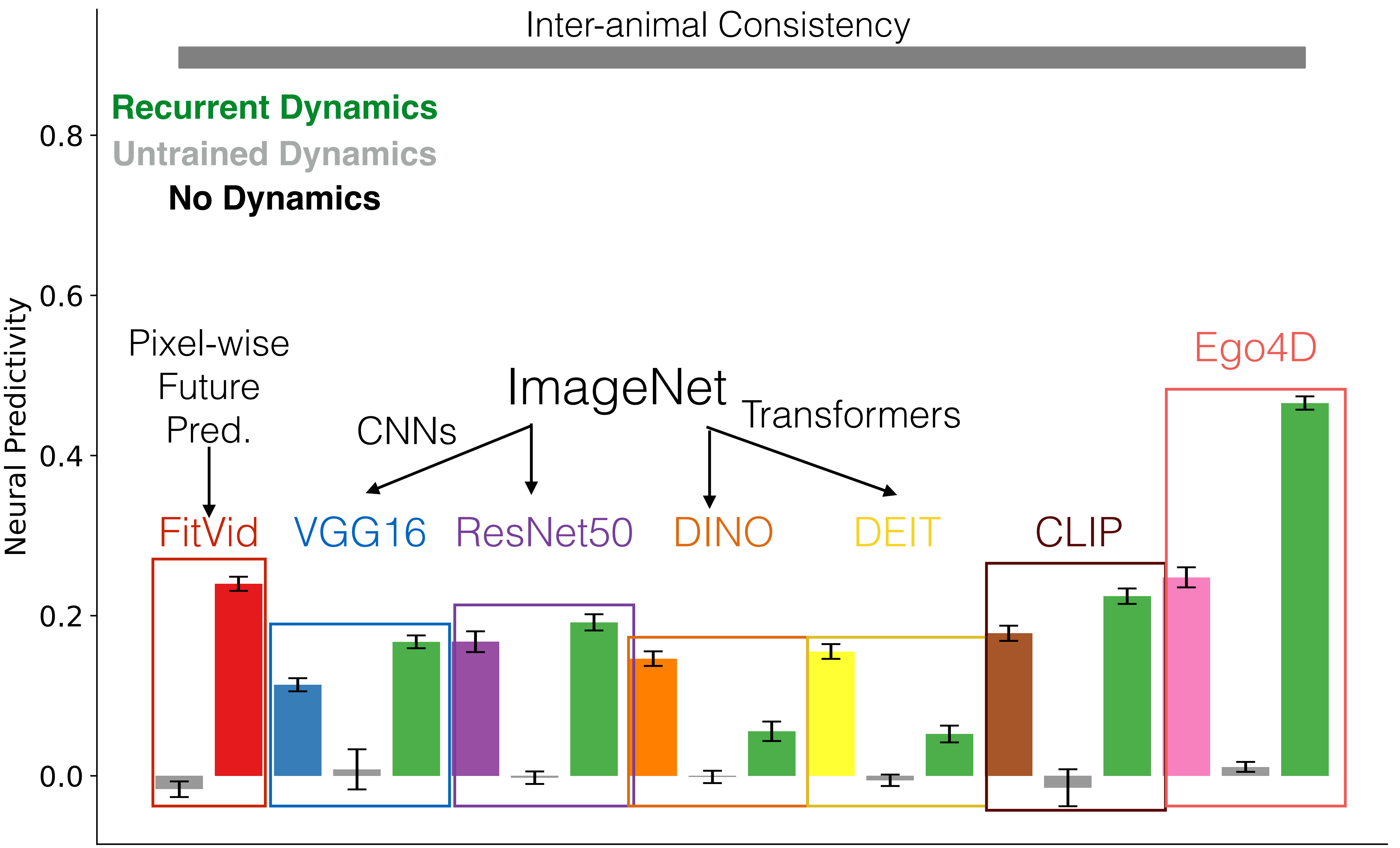



Rishi Rajalingham

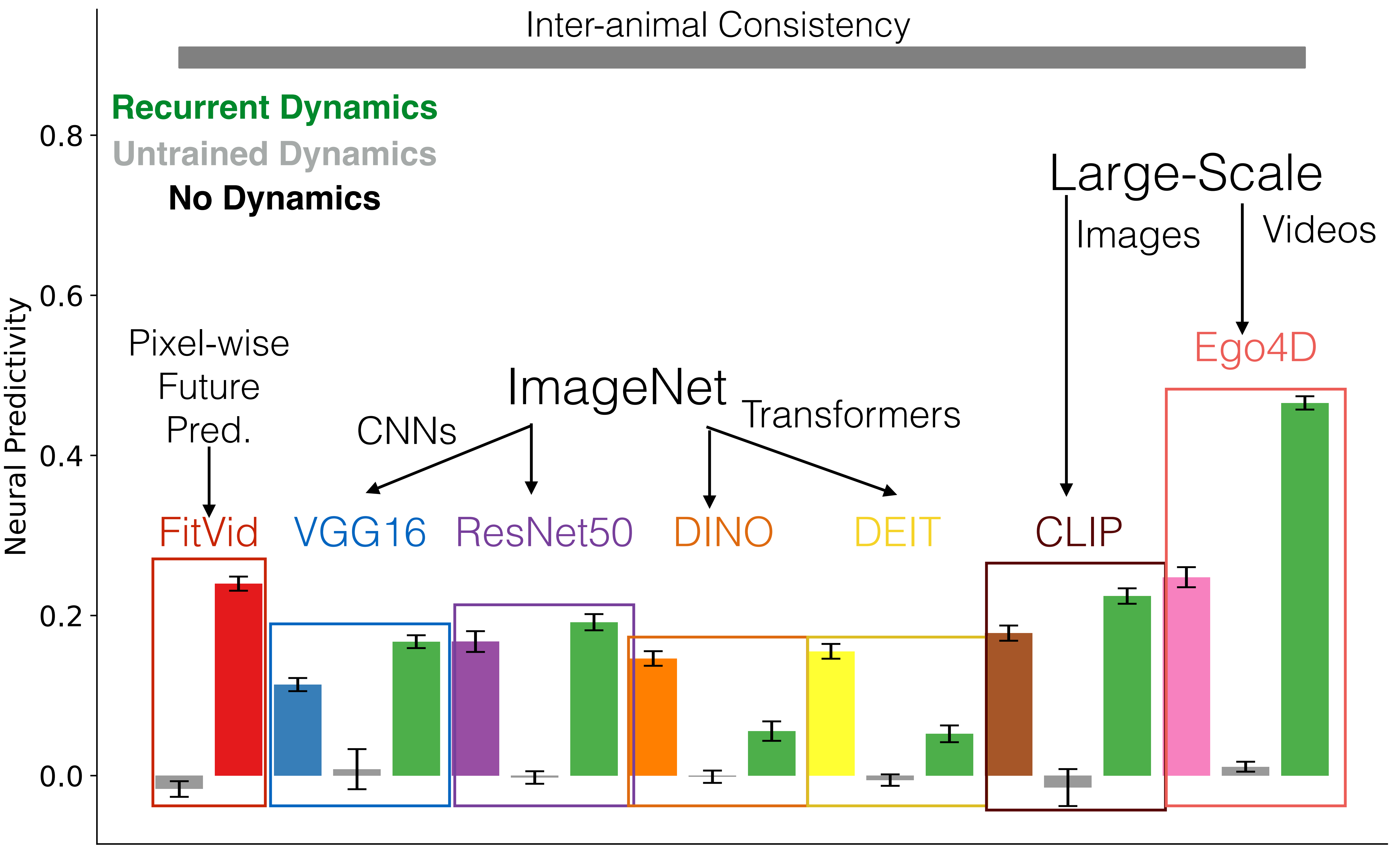
Preliminary Results Modeling DMFC



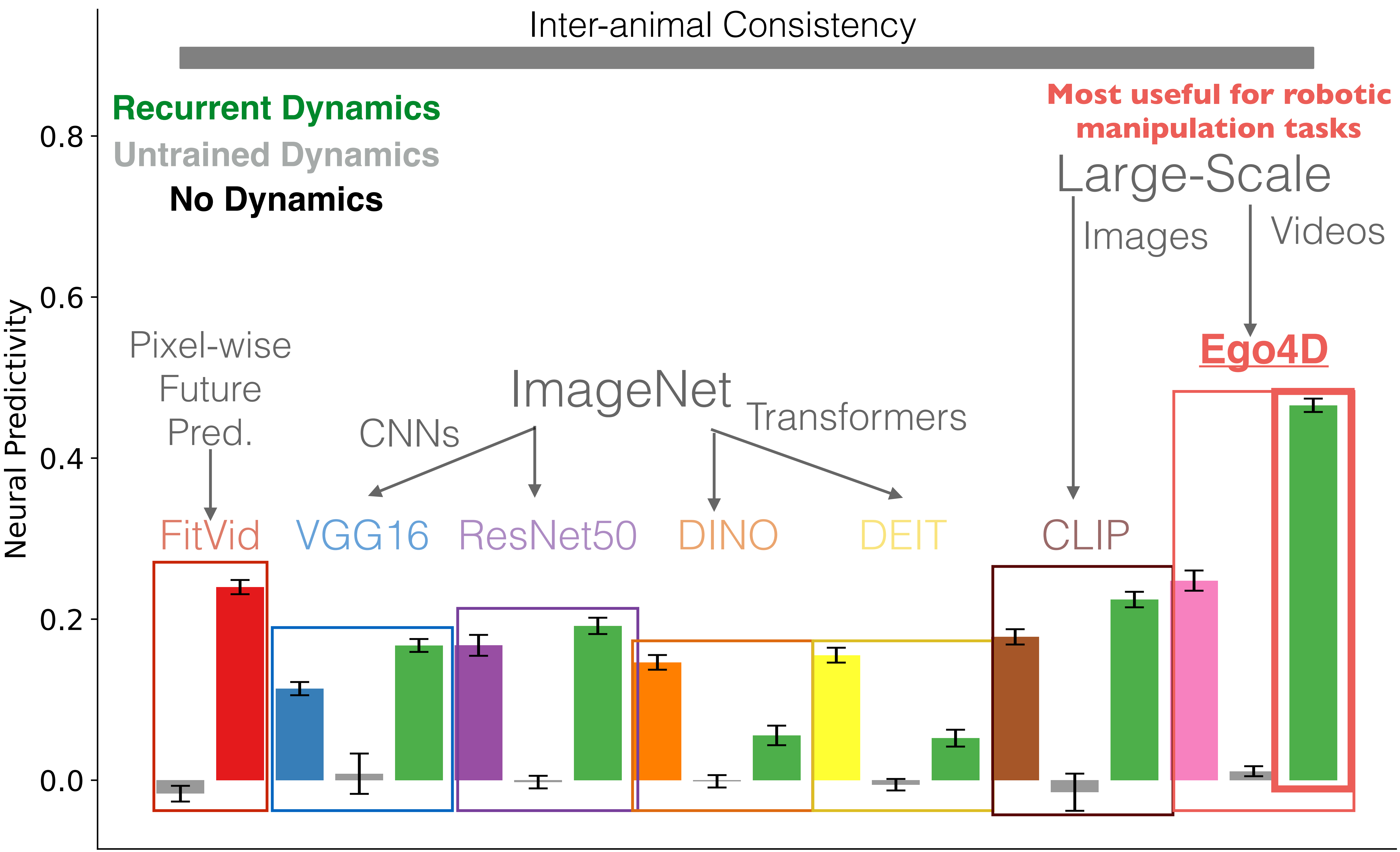
Preliminary Results Modeling DMFC



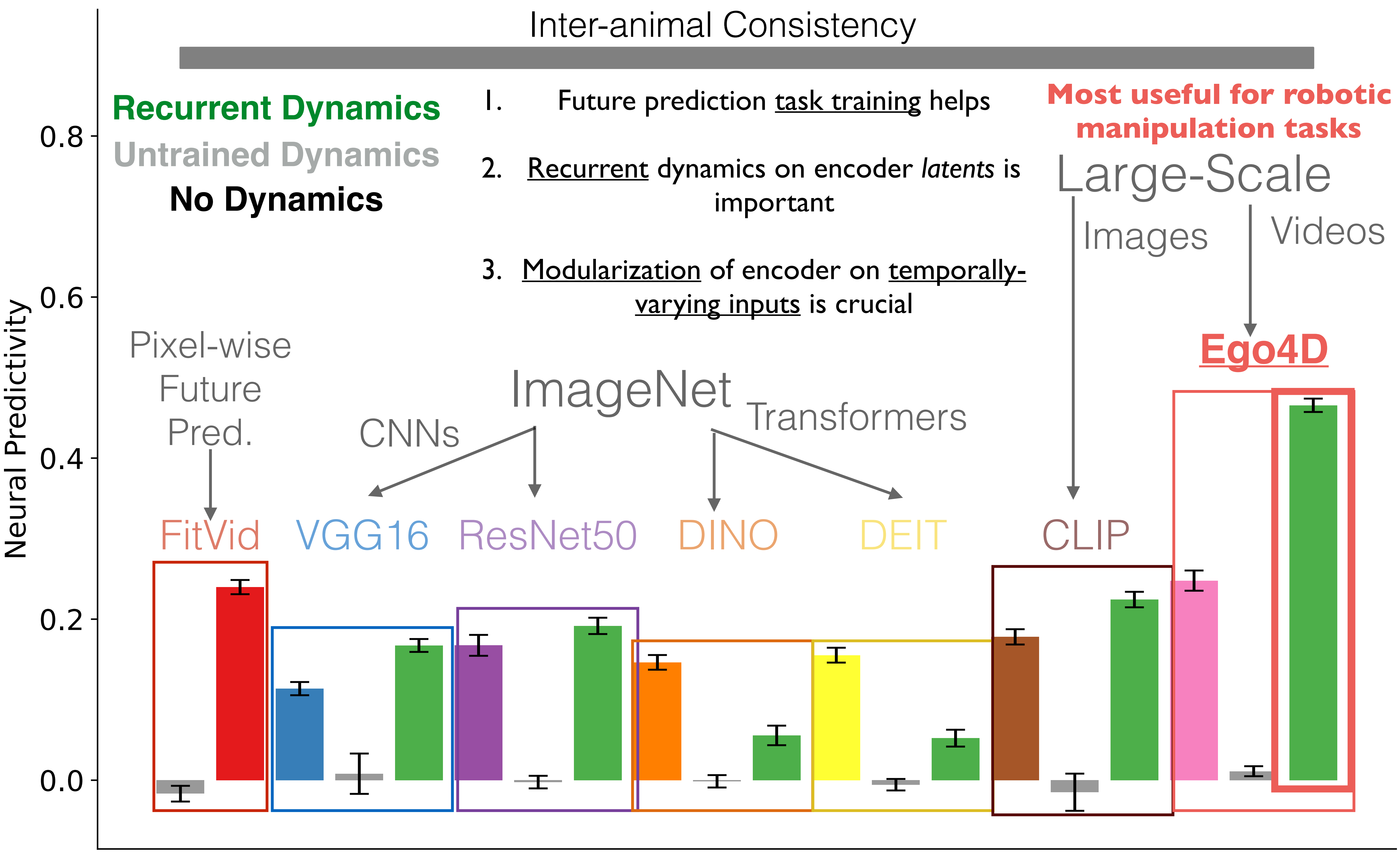
Preliminary Results Modeling DMFC



Preliminary Results Modeling DMFC

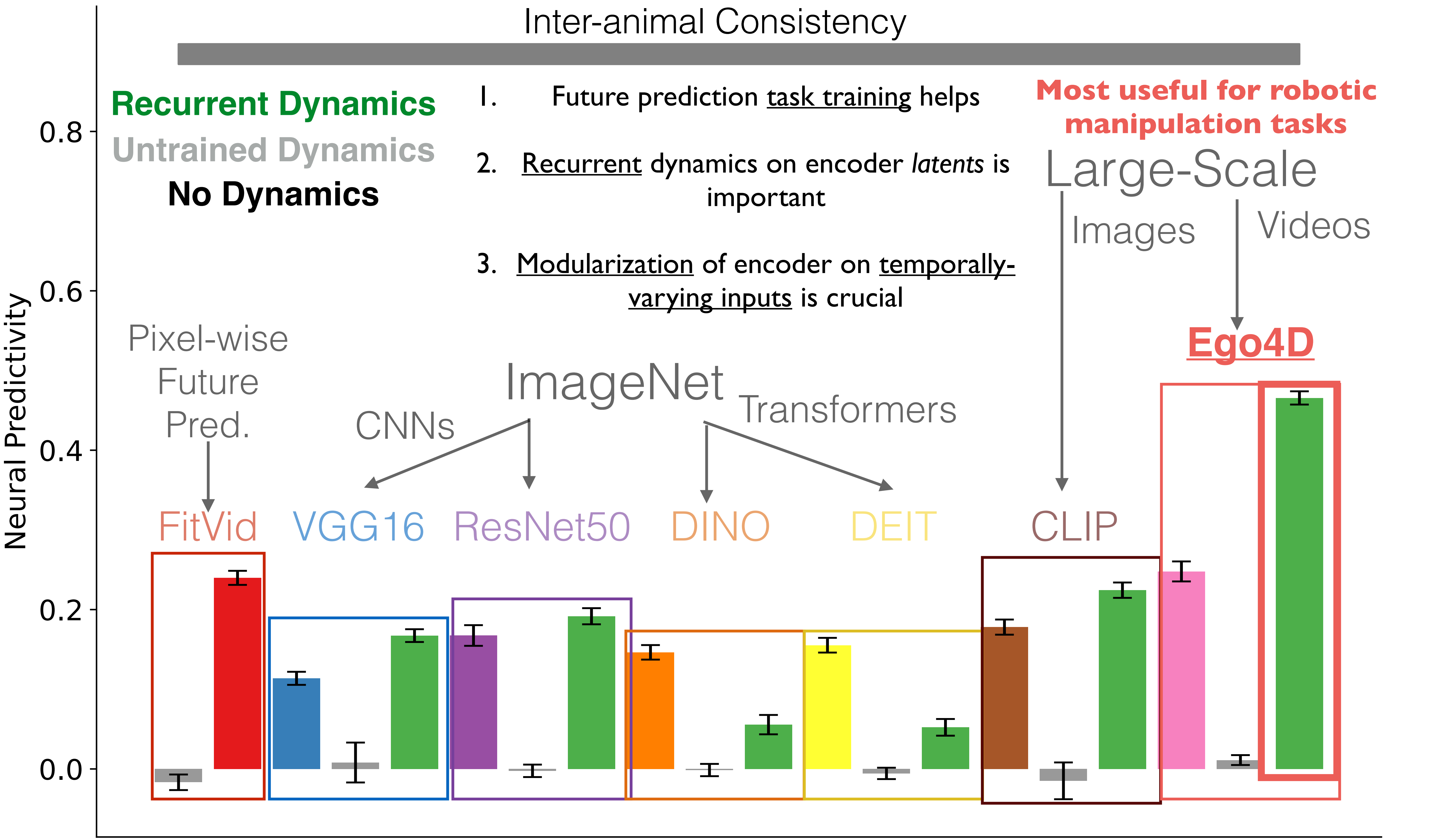


Preliminary Results Modeling DMFC

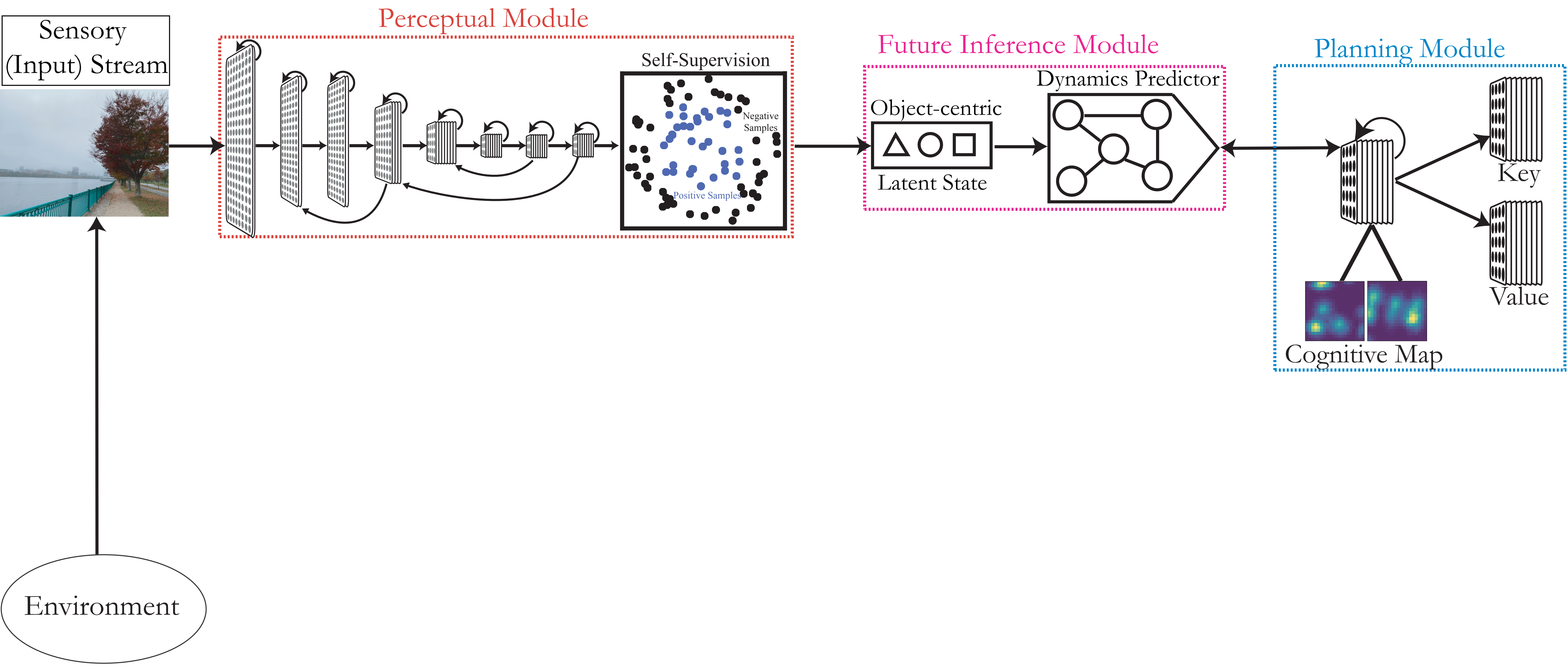


Preliminary Results Modeling DMFC

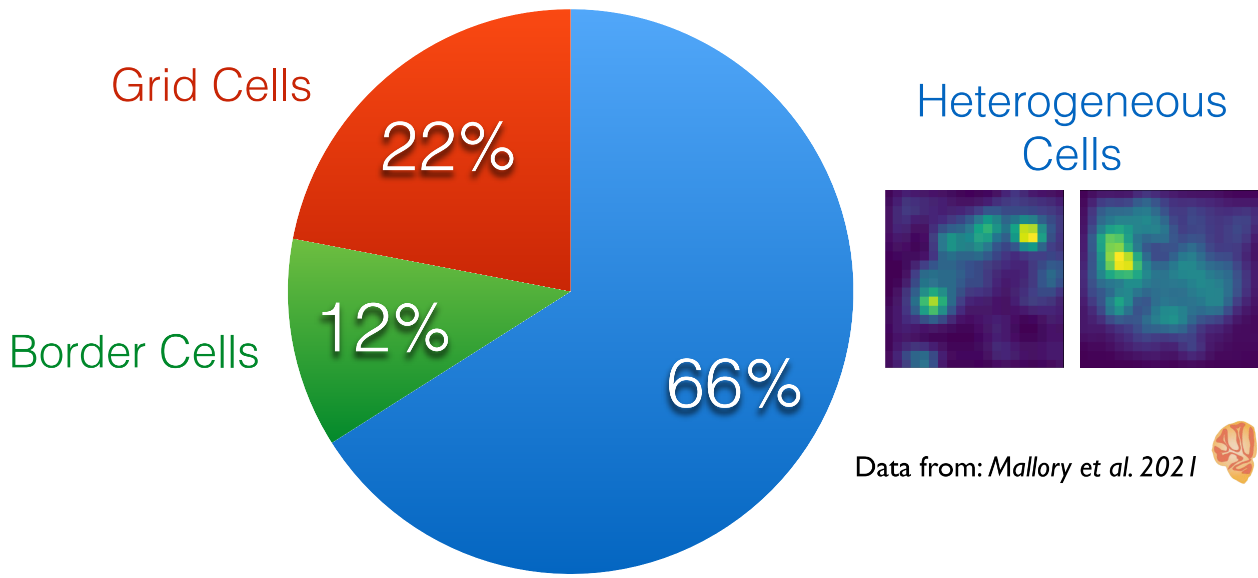
DMFC appears to be optimized to future predict on *dynamic*, *scene-centric* representations



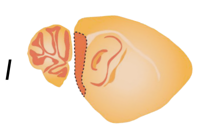
Next Steps: Cognitive Maps to Enable Navigation and Long-Range Planning



Next Steps: Cognitive Maps to Enable Navigation and Long-Range Planning



Data from: *Mallory et al.* 2021



<u>Neurobiological Puzzle(s):</u>

1. How might we characterize what these heterogeneous cells do?

2. What functional role do these cells serve in the circuit, if any?

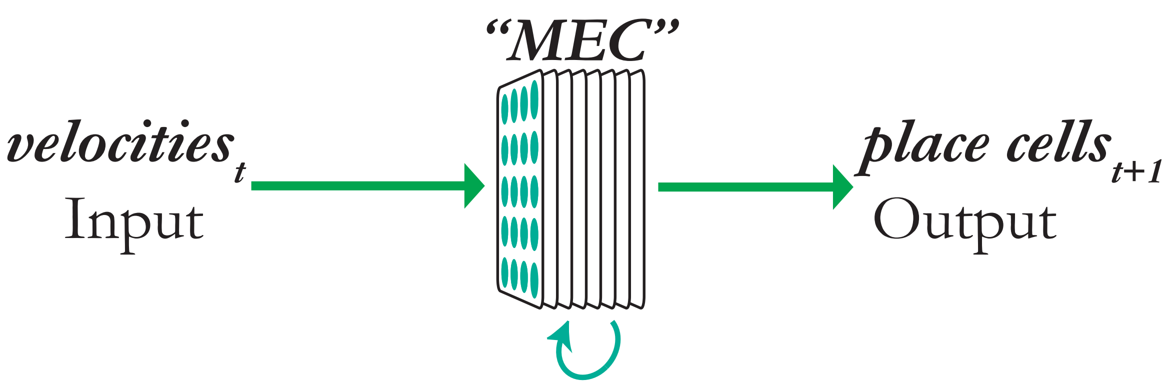
<u>*Partial* Resolution:</u>

1. <u>Characterization:</u> Close to perfect neural predictivity with the above constraints — more complex environments are needed!

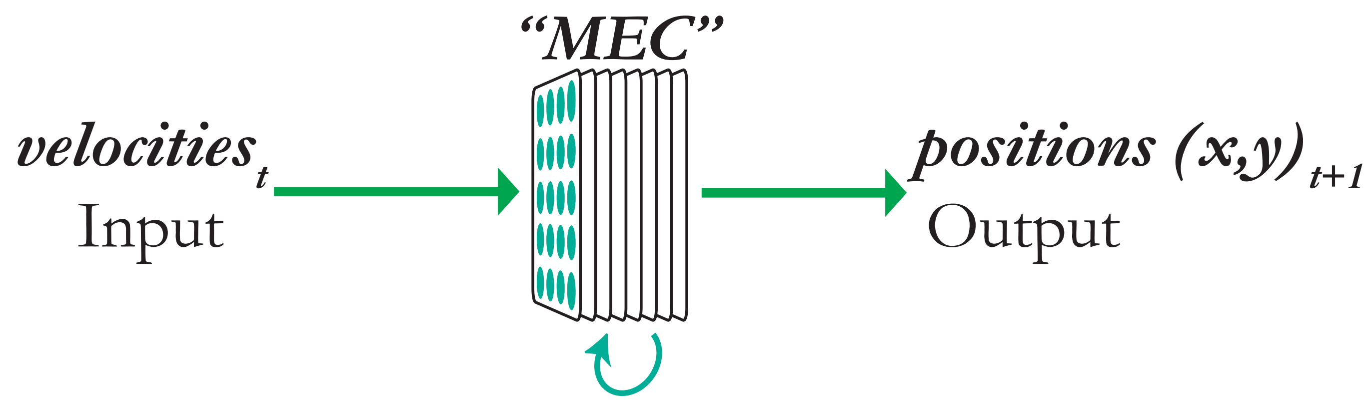
2. <u>Functional Role:</u> Grid cells are not functionally unique! Both heterogeneous and grid cells arise jointly through task optimization.

Next Steps: Cognitive Maps to Enable Navigation and Long-Range Planning

$$\mathcal{L}(\hat{p}, p) := -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N_p} p_i^t \log \hat{p}_i^t$$



$$\mathcal{L}(\hat{p}, p) := \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left(\left(p_x^t - \widehat{p_x^t}\right)^2 + \left(p_y^t - \widehat{p_y^t}\right)^2\right)$$



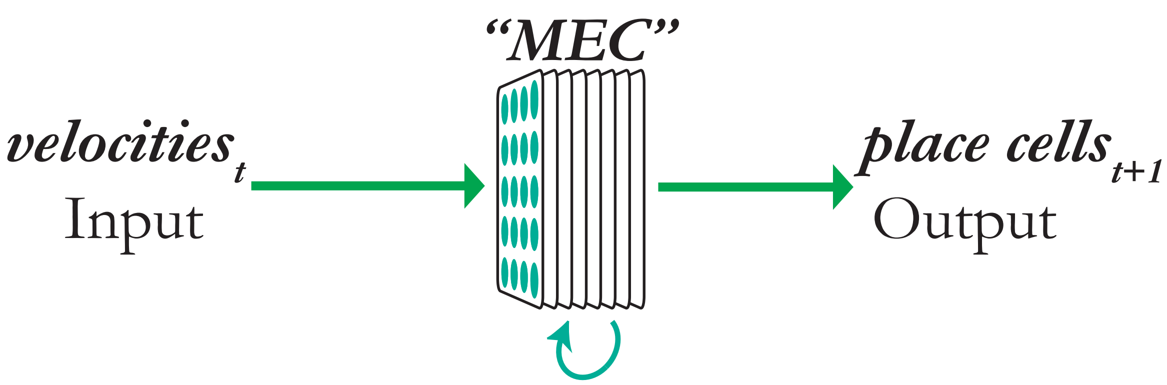
A. Nayebi, *et al.*
Explaining Heterogeneity in Medial Entorhinal Cortex with Task-Driven Neural Networks.
NeurIPS 2021 (spotlight)

<u>Core Unsolved Question:</u>

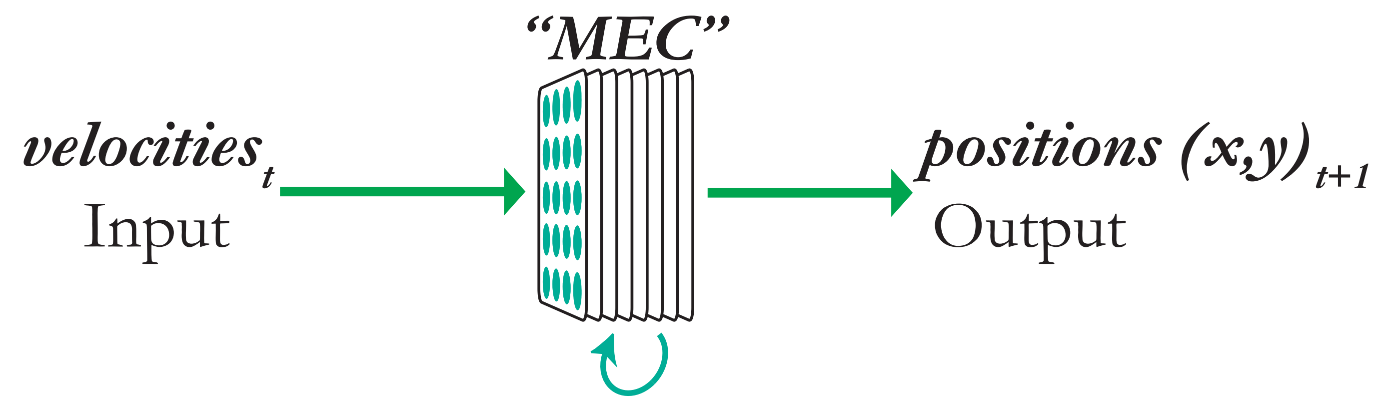
Direct path integration fails to match MEC, intermediate place cell representation is needed. For what normative purpose?

Next Steps: Cognitive Maps to Enable Navigation and Long-Range Planning

$$\mathcal{L}(\hat{p}, p) := -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N_p} p_i^t \log \hat{p}_i^t$$



$$\mathcal{L}(\hat{p}, p) := \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left(\left(p_x^t - \widehat{p_x^t}\right)^2 + \left(p_y^t - \widehat{p_y^t}\right)^2\right)$$



A. Nayebi, *et al.*
Explaining Heterogeneity in Medial Entorhinal Cortex with Task-Driven Neural Networks.
NeurIPS 2021 (spotlight)

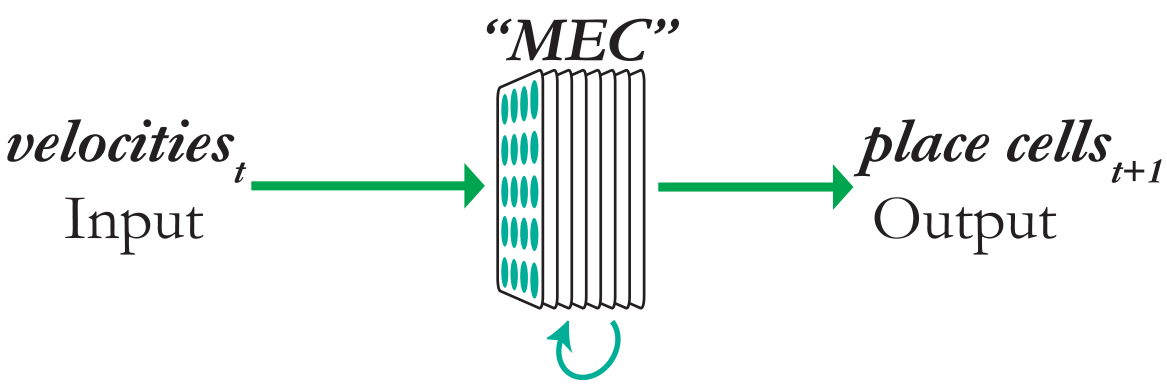
Core Unsolved Question:

Direct path integration fails to match MEC, intermediate place cell representation is needed. For what normative purpose?

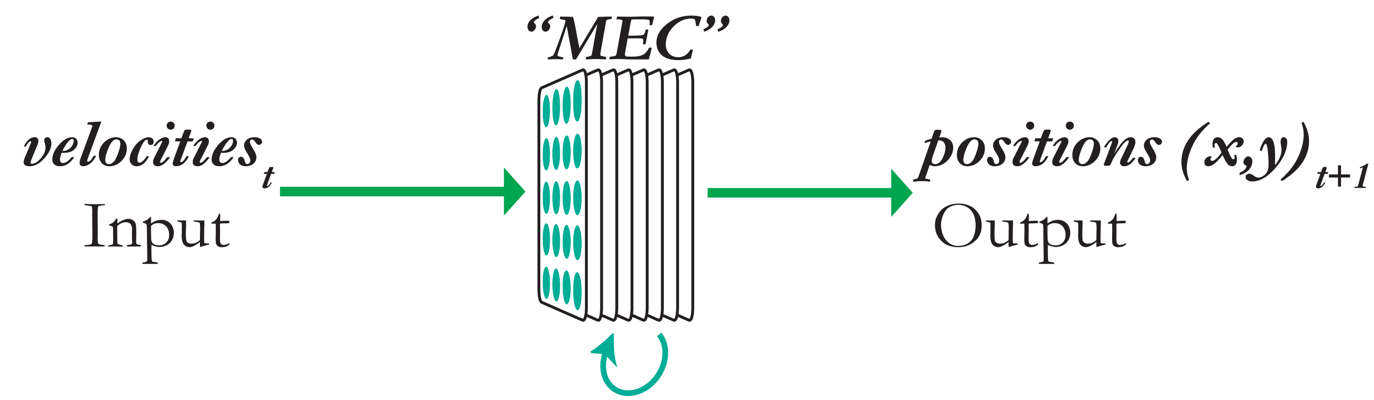
Can this coordination help enable long-range planning, despite intervening events over long time horizons?

Next Steps: Cognitive Maps to Enable Navigation and Long-Range Planning

$$\mathcal{L}(\hat{p}, p) := -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N_p} p_i^t \log \hat{p}_i^t$$



$$\mathcal{L}(\hat{p}, p) := \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left(\left(p_x^t - \widehat{p_x^t}\right)^2 + \left(p_y^t - \widehat{p_y^t}\right)^2\right)$$



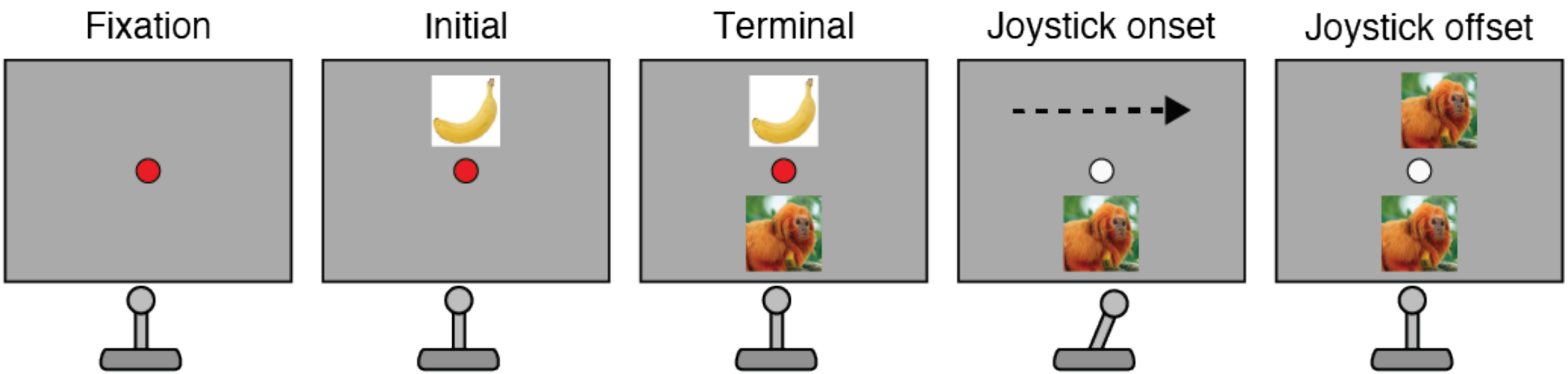
A. Nayebi, *et al.*
Explaining Heterogeneity in Medial Entorhinal Cortex with Task-Driven Neural Networks.
NeurIPS 2021 (spotlight)

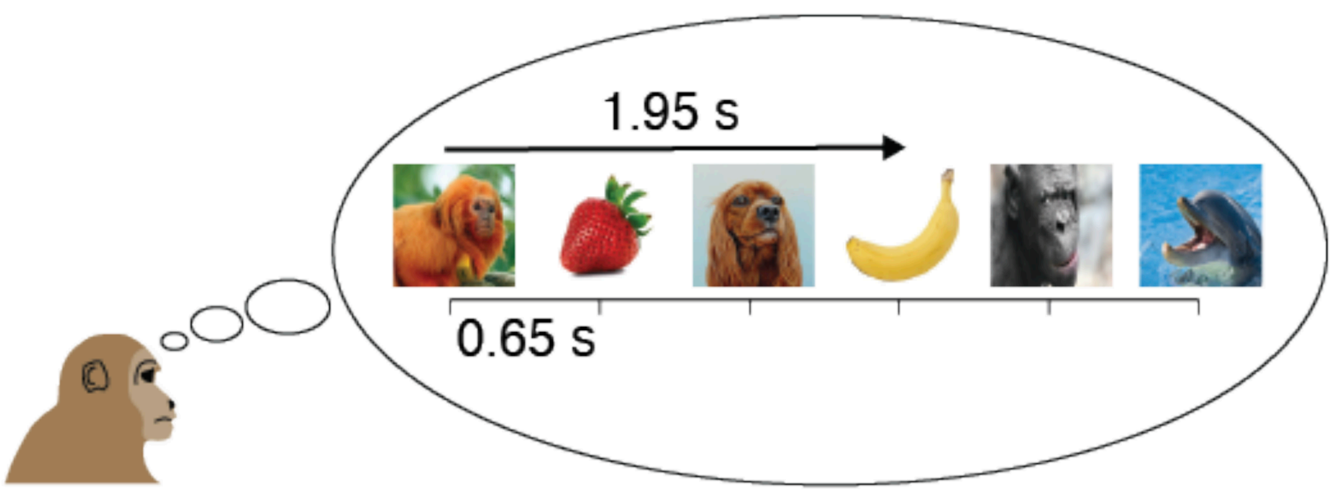
<u>*Core Unsolved Question:*</u>

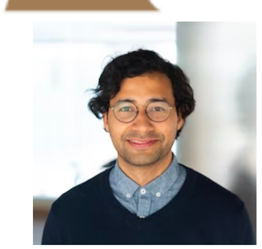
Direct path integration fails to match MEC, intermediate place cell representation is needed. For what normative purpose?

Can this coordination help enable long-range planning, despite <u>intervening events</u> over <u>long time horizons</u>?

Next Steps: Cognitive Maps to Enable Navigation and Long-Range Planning



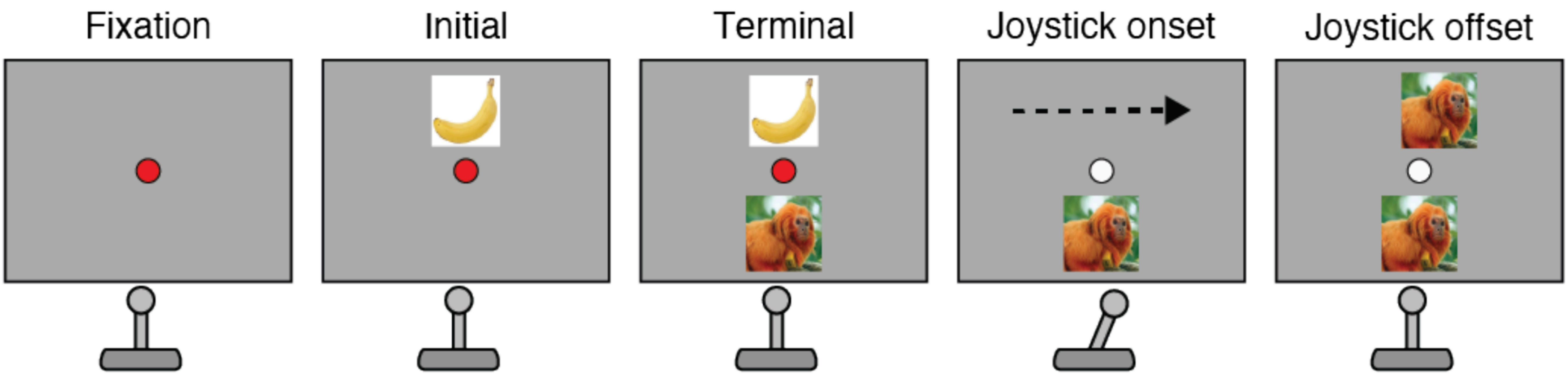


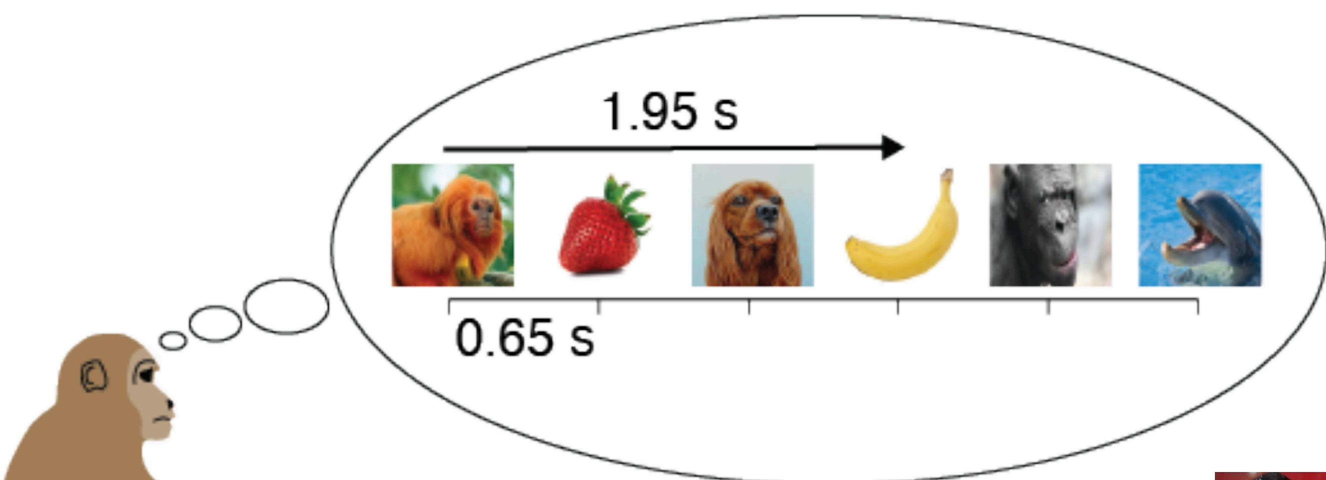

Sujaya Neupane

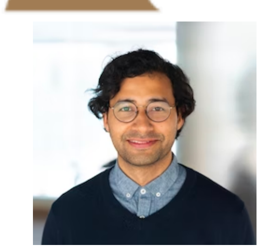

Mehrdad Jazayeri

Can this coordination help enable long-range planning, despite intervening events over long time horizons?

Next Steps: Cognitive Maps to Enable Navigation and Long-Range Planning







Sujaya Neupane

Mehrdad Jazayeri

Paired Associative Inference (PAI) Task:

A B C

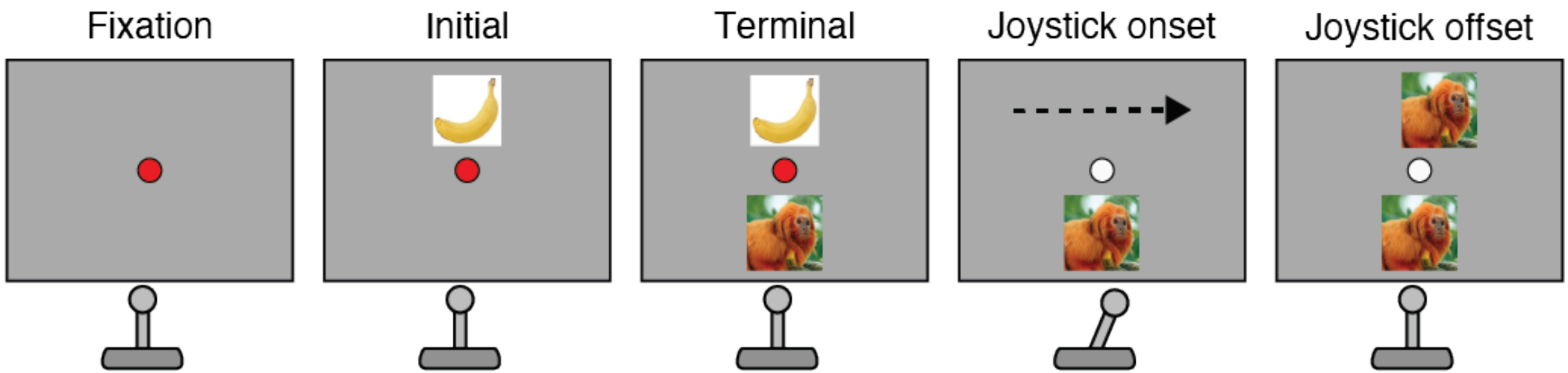
Two types of queries:

A “Direct” B

A “Indirect” C

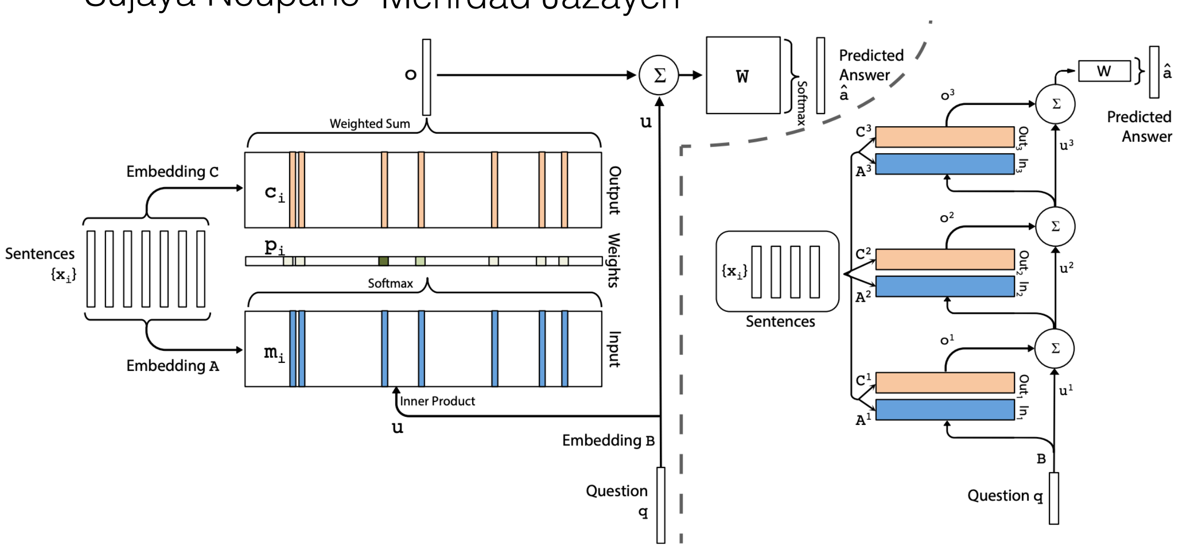
Can this coordination help enable long-range planning, despite intervening events over long time horizons?

Next Steps: Cognitive Maps to Enable Navigation and Long-Range Planning

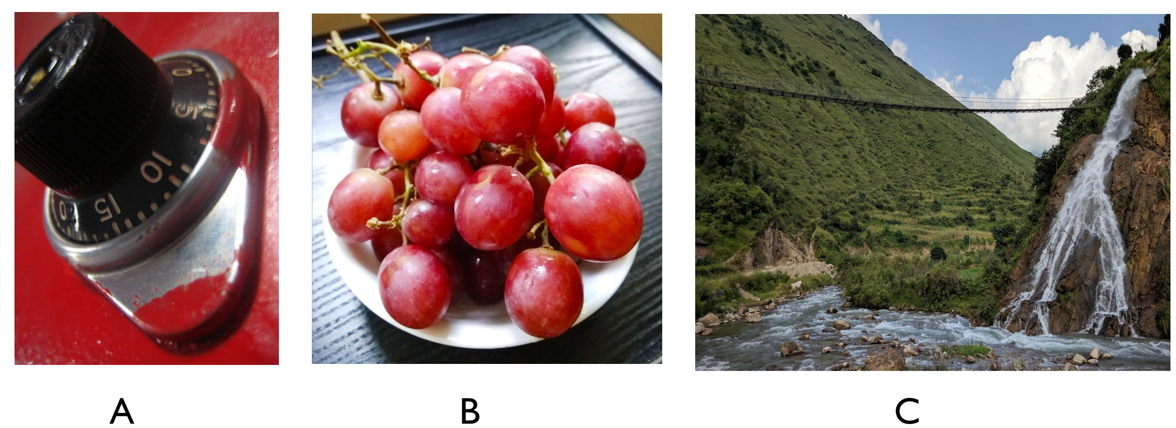



Sujaya Neupane

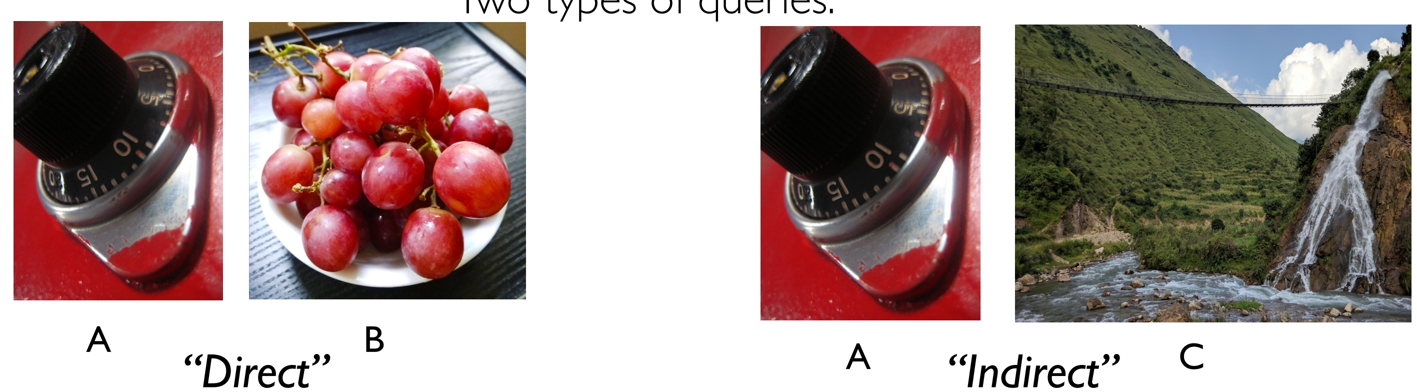

Mehrdad Jazayeri



Paired Associative Inference (PAI) Task:

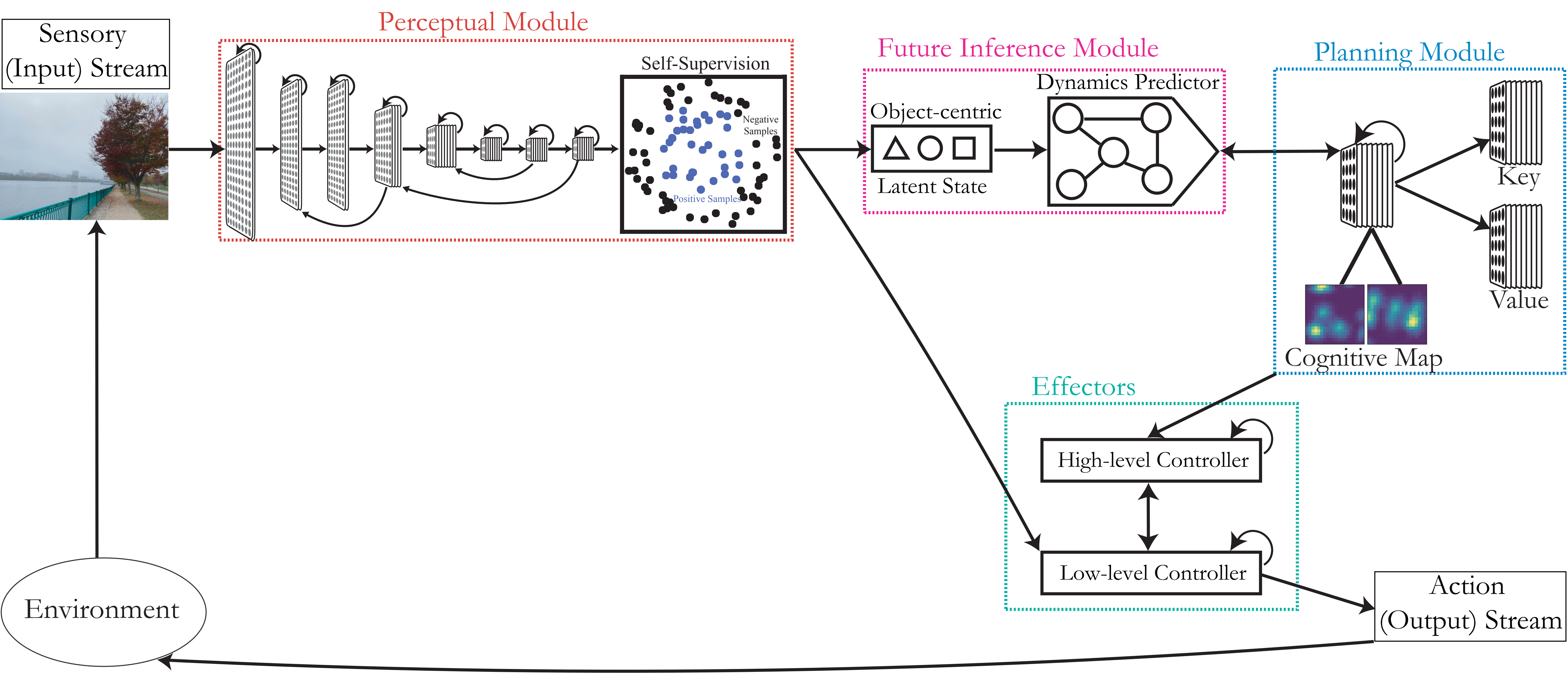


Two types of queries:

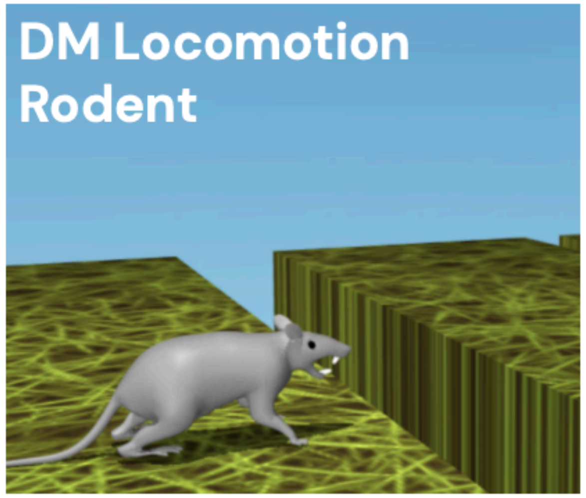


Can this coordination help enable long-range planning, despite <u>intervening events</u> over <u>long time horizons</u>?

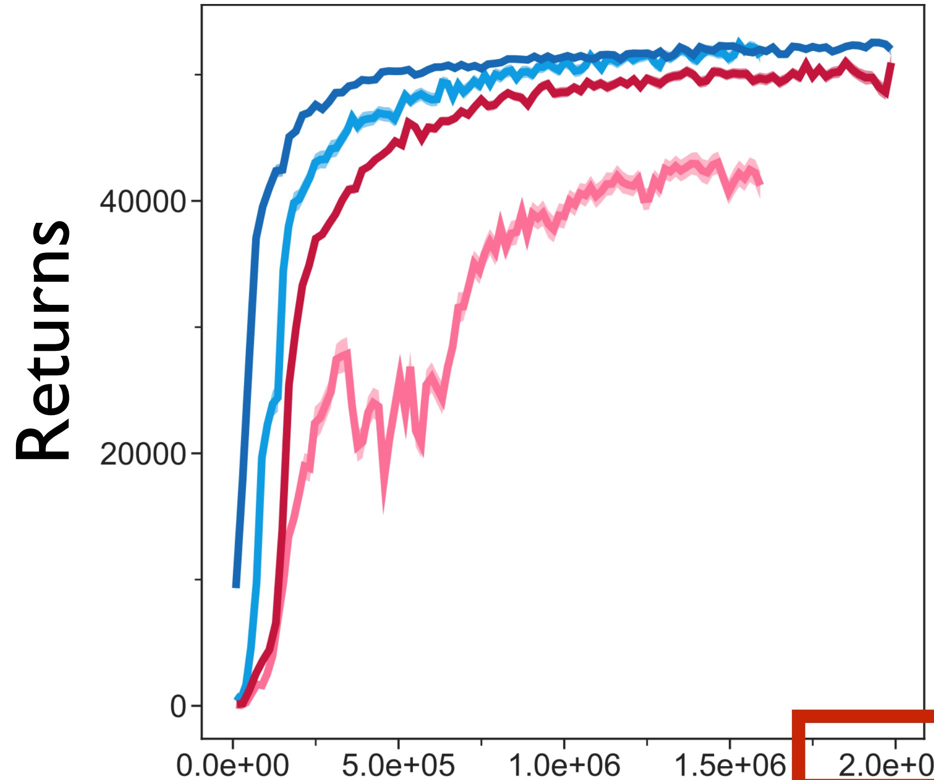
Next Steps: Sensorimotor Coordination in Ethologically-Relevant Environments



Model-free Control of Animal-like Bodies is Hard

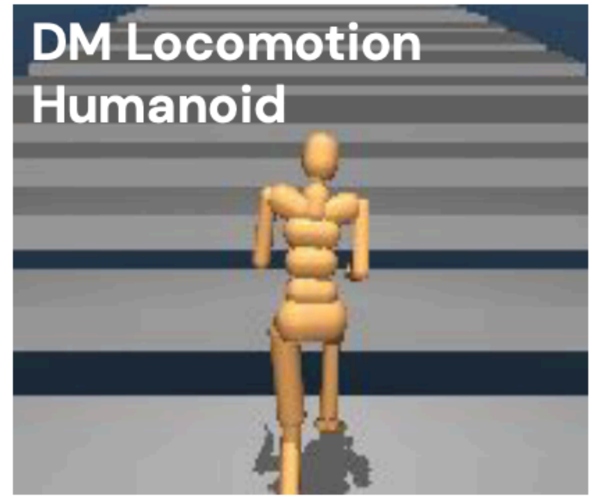
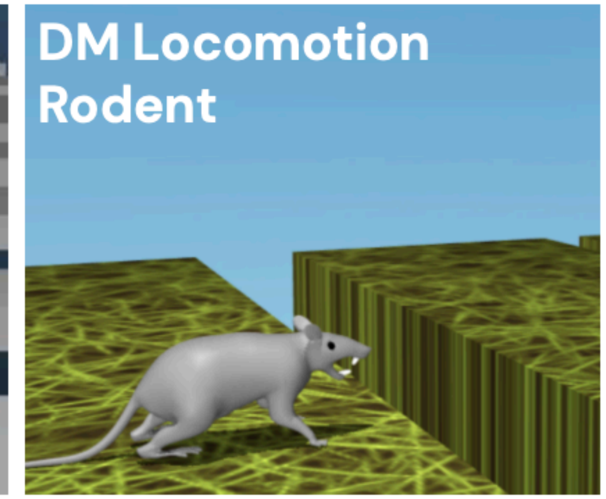
Task domain			
Action space	continuous	continuous	continuous
Observation space	state	pixels	pixels
Exploration difficulty	low to moderate	high	moderate
Dynamics	deterministic / stochasitic	deterministic	deterministic

Merel, Aldorando*, Marshal* et al. 2020; Gulcehre et al. 2021*

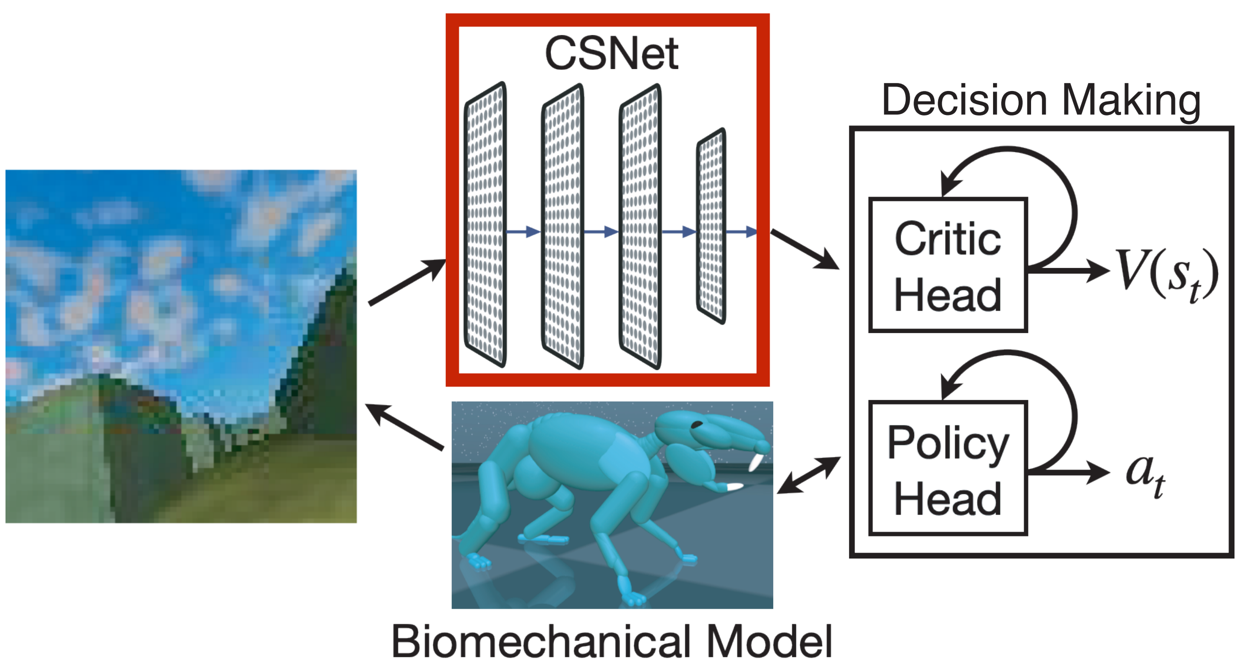


Millions of iterations from scratch!

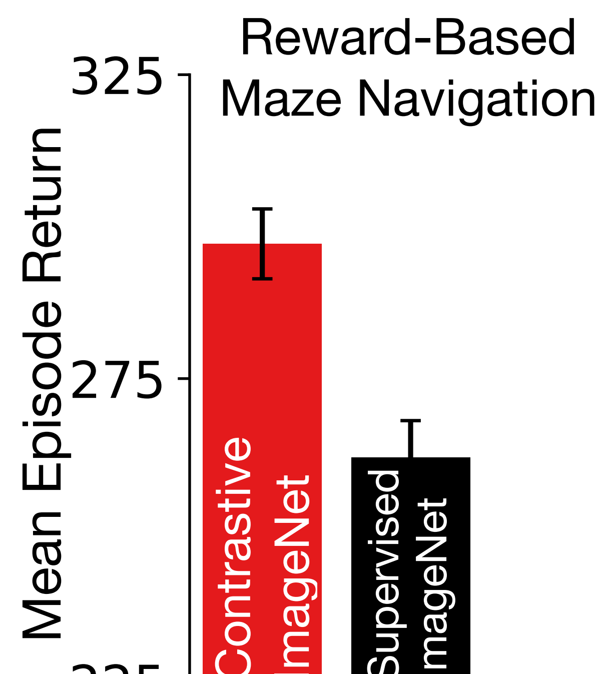
Self-supervised pre-training of encoders seems to help

Task domain	DM Control Suite / Real World RL Suite	DM Locomotion Humanoid	DM Locomotion Rodent
			
Action space	continuous	continuous	continuous
Observation space	state	pixels	pixels
Exploration difficulty	low to moderate	high	moderate
Dynamics	deterministic / stochasitic	deterministic	deterministic

Merel, Aldorando*, Marshal* et al. 2020; Gulcehre et al. 2021*



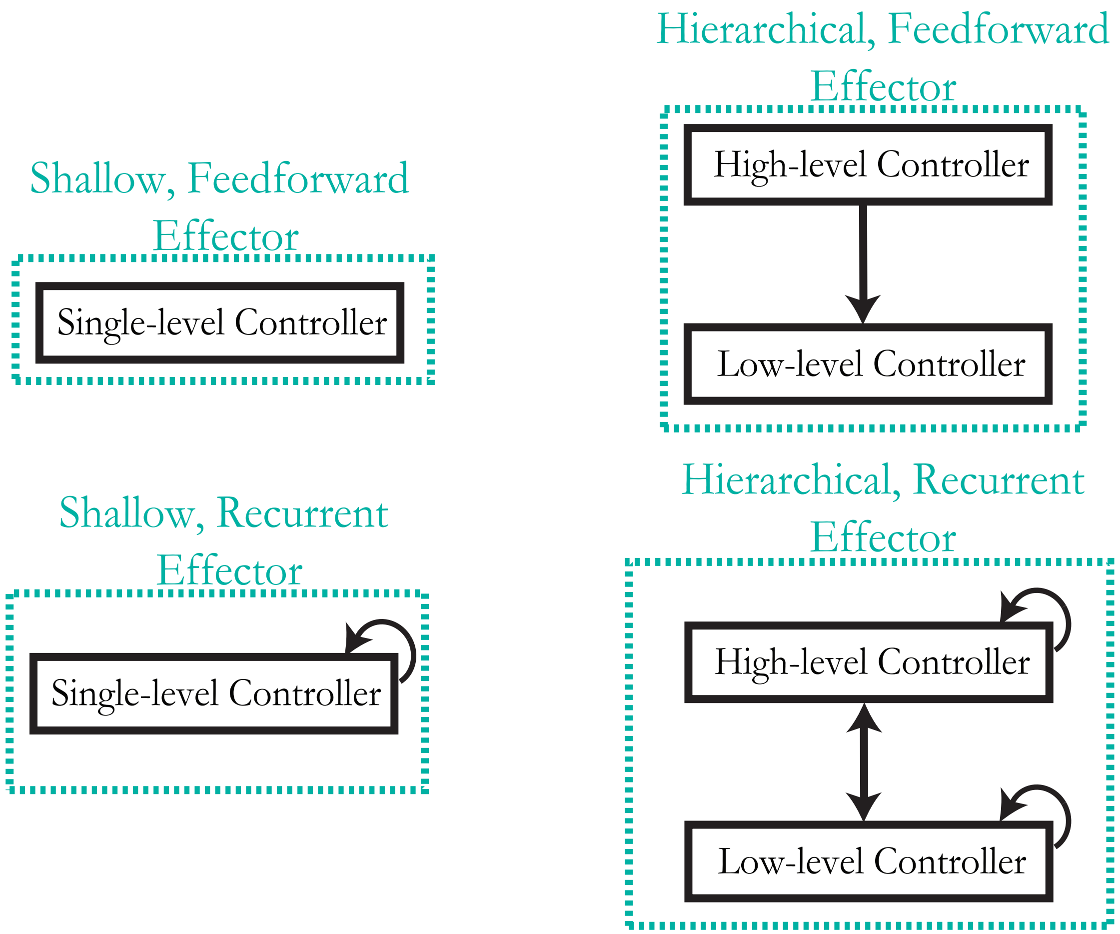
(*Merel*, Aldorando*, Marshal* et al. 2020*)



Nayebi, Kong* et al. 2022*

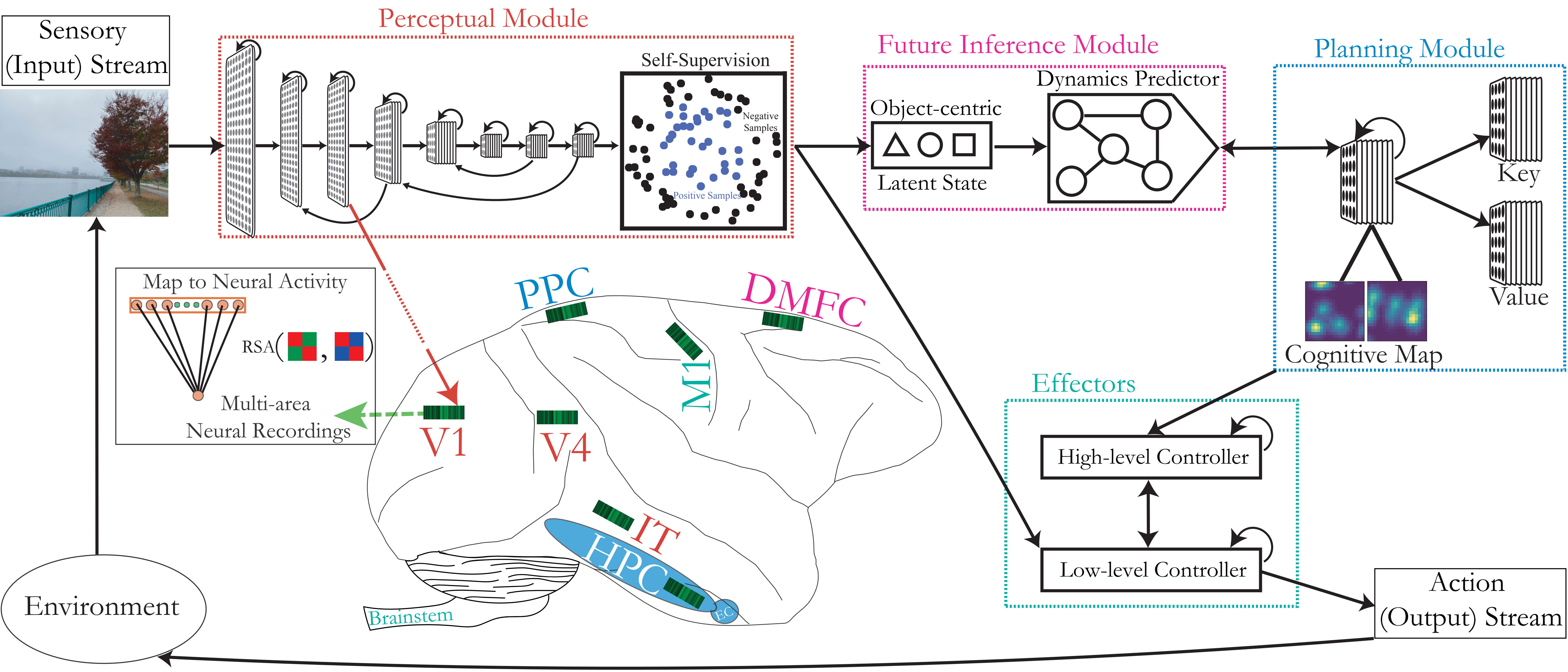
Effector architectures

Deitke et al. 2022



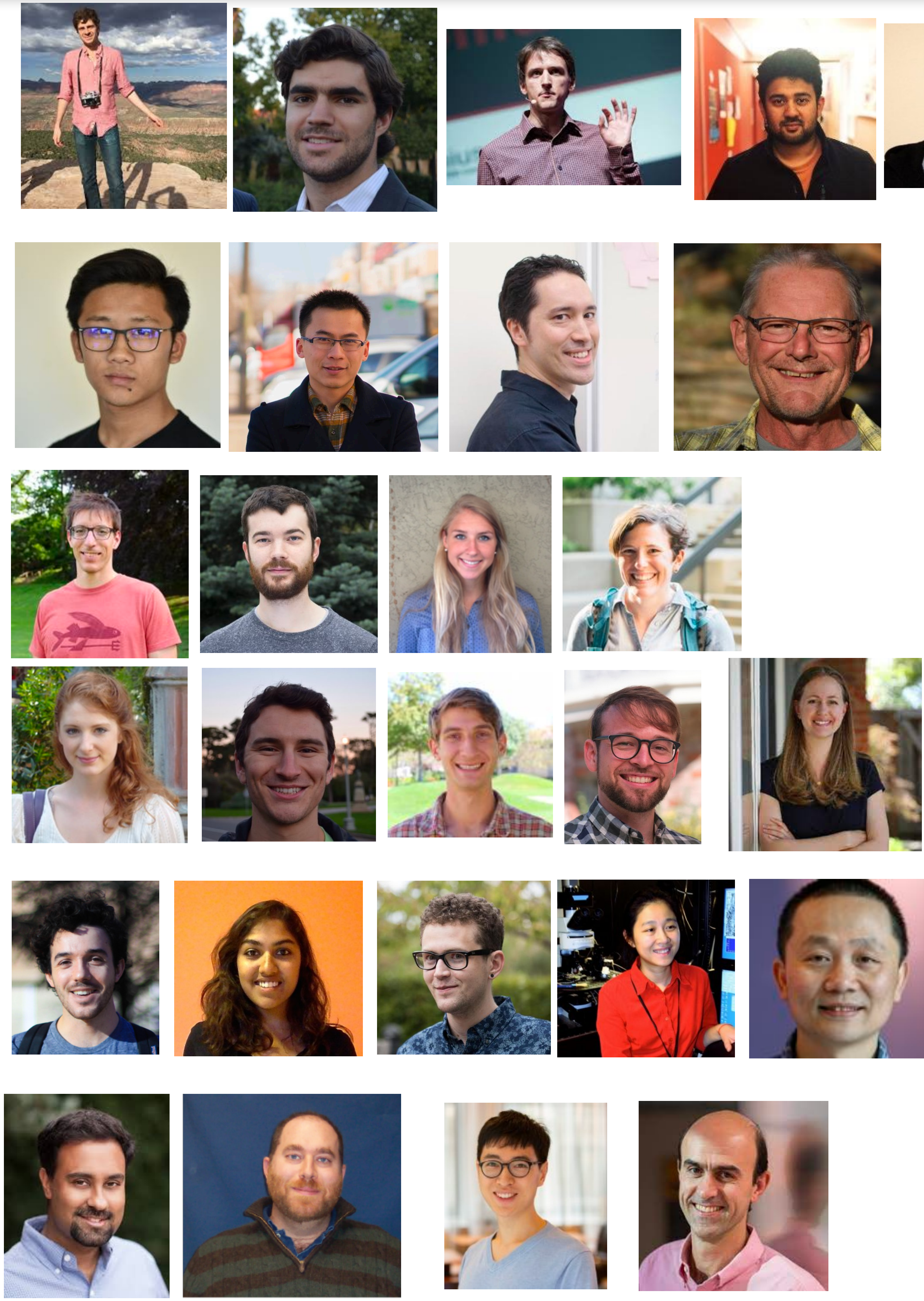
Normative basis of motor coordination (e.g. involving the interaction of higher motor cortices, basal ganglia, and the brainstem)

Integrative Agents to Reverse-Engineer Natural Cognition



Building integrative agents in rich, ethologically-relevant environments will be the basis for the evolutionary design principles of *natural cognition*

Acknowledgements



Contact:
anayebi@mit.edu

@aran_nayebi

Funding:

K. Lisa Yang ICoN Postdoctoral Fellowship, McGovern Institute, MIT

Stanford Neurosciences PhD Program

Stanford Mind, Brain, Computation and Technology Training Program, Wu Tsai Neurosciences Institute