CS375 / Psych 249:
Large-Scale Neural Network Models for Neuroscience

Lecture 7: Recurrence and Feedback in the Visual System

2025.02.01

Daniel Yamins
Departments of Computer Science and of Psychology
Stanford Neuroscience and Artificial Intelligence Laboratory
Wu Tsai Neurosciences Institute
Stanford University

Four Principles of Goal-Driven Modeling

1.

A = *architecture class*

2.

T = *task/objective*

3.

D = *dataset*

4.

L = *learning rule*

Four Principles of Goal-Driven Modeling

1.

A = *architecture class*

2.

T = *task/objective*

3.

D = *dataset*

4.

L = *learning rule*

Best proxies thus far for ventral stream:

A = *ConvNets of reasonable depth*

T = *multi-way object categorization*

D = *ImageNet images*

L = *evolutionary architecture search + filter learning through gradient descent*

Four Principles of Goal-Driven Modeling

1.

A = *architecture class* ***= circuit neuro-anatomy***

2.

T = *task/objective* ***= ecological niche***

3.

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

4.

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

Best proxies thus far for ventral stream:

A = *ConvNets of reasonable depth*

T = *multi-way object categorization*

D = *ImageNet images*

L = *evolutionary architecture search + filter learning through gradient descent*

Four Principles of Goal-Driven Modeling

1.

A = *architecture class* **= *circuit neuro-anatomy***

solving

2.

T = *task/objective* **= *ecological niche***

situated in

3.

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

updating according to

4.

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

Best proxies thus far for ventral stream:

A = *ConvNets of reasonable depth*

T = *multi-way object categorization*

D = *ImageNet images*

L = *evolutionary architecture search + filter learning through gradient descent*

Big Problems in Each Area

***bad** = obviously deeply wrong as model of the brain or behavior

PROBLEM

1. ✗ ***bad***

A = *architecture class*

e.g. **CNNs**

2.

T = *task/objective*

e.g. **Object Categorization**

3.

D = *dataset*

e.g. **ImageNet**

4.

L = *learning rule*

e.g. **Arch. Srch.** + **Grad. Desc.**

Big Problems in Each Area

***bad** = obviously deeply wrong as model of the brain or behavior

1. ✗ ***bad***

A = *architecture class*

e.g. **CNNs**

2.

T = *task/objective*

e.g. **Object Categorization**

3.

D = *dataset*

e.g. **ImageNet**

4.

L = *learning rule*

e.g. **Arch. Srch.** + **Grad. Desc.**

PROBLEM

RECURRENCE and FEEDBACK!!?

Big Problems in Each Area

***bad** = obviously deeply wrong as model of the brain or behavior

PROBLEM

1. ✗ ***bad***

A = *architecture class*

e.g. **CNNs**

RECURRENCE and FEEDBACK!!?

2. ✗ ***bad***

T = *task/objective*

e.g. **Object Categorization**

TOO MUCH LABELLED DATA REQUIRED!!?

3.

D = *dataset*

e.g. **ImageNet**

4.

L = *learning rule*

e.g. **Arch. Srch. + Grad. Desc.**

Big Problems in Each Area

***bad** = obviously deeply wrong as model of the brain or behavior

PROBLEM

1. ❌ ***bad***

A = *architecture class*

e.g. **CNNs**

RECURRENCE and FEEDBACK!!?

2. ❌ ***bad***

T = *task/objective*

e.g. **Object Categorization**

TOO MUCH LABELLED DATA REQUIRED!!?

3. ❌ ***bad***

D = *dataset*

e.g. **ImageNet**

REAL NOISY VIDEO DATASTREAMS vs STEREOTYPED CLEAN STILL IMAGES

4.

L = *learning rule*

e.g. **Arch. Srch.** + **Grad. Desc.**

Big Problems in Each Area

***bad** = obviously deeply wrong as model of the brain or behavior

PROBLEM

1. ✗ **bad**

***A** = architecture class*

e.g. **CNNs**

RECURRENCE and FEEDBACK!!?

2. ✗ **bad**

***T** = task/objective*

e.g. **Object Categorization**

TOO MUCH LABELLED DATA REQUIRED!!?

3. ✗ **bad**

***D** = dataset*

e.g. **ImageNet**

REAL NOISY VIDEO DATASTREAMS vs STEREOTYPED CLEAN STILL IMAGES

4. ✗ **bad**

***L** = learning rule*

e.g. **Arch. Srch. + Grad. Desc.**

BACKPROP AND ITS DISCONTENTS

From Last Time . . .

*✓**ok** = we've really nailed it
*✓**ok-ish** = **harder to reject out of hand**
***bad** = obviously deeply wrong

1. ✗ **bad**

***A** = architecture class*

ConvRNNs

2. ✓**ok**

***T** = task/objective*

e.g. **Object Categorization**

3. *✓**ok-ish**

***D** = dataset*

e.g. **ImageNet**

4. ✗ **bad**

***L** = learning rule*

e.g. **Arch. Srch. + Grad. Desc.**

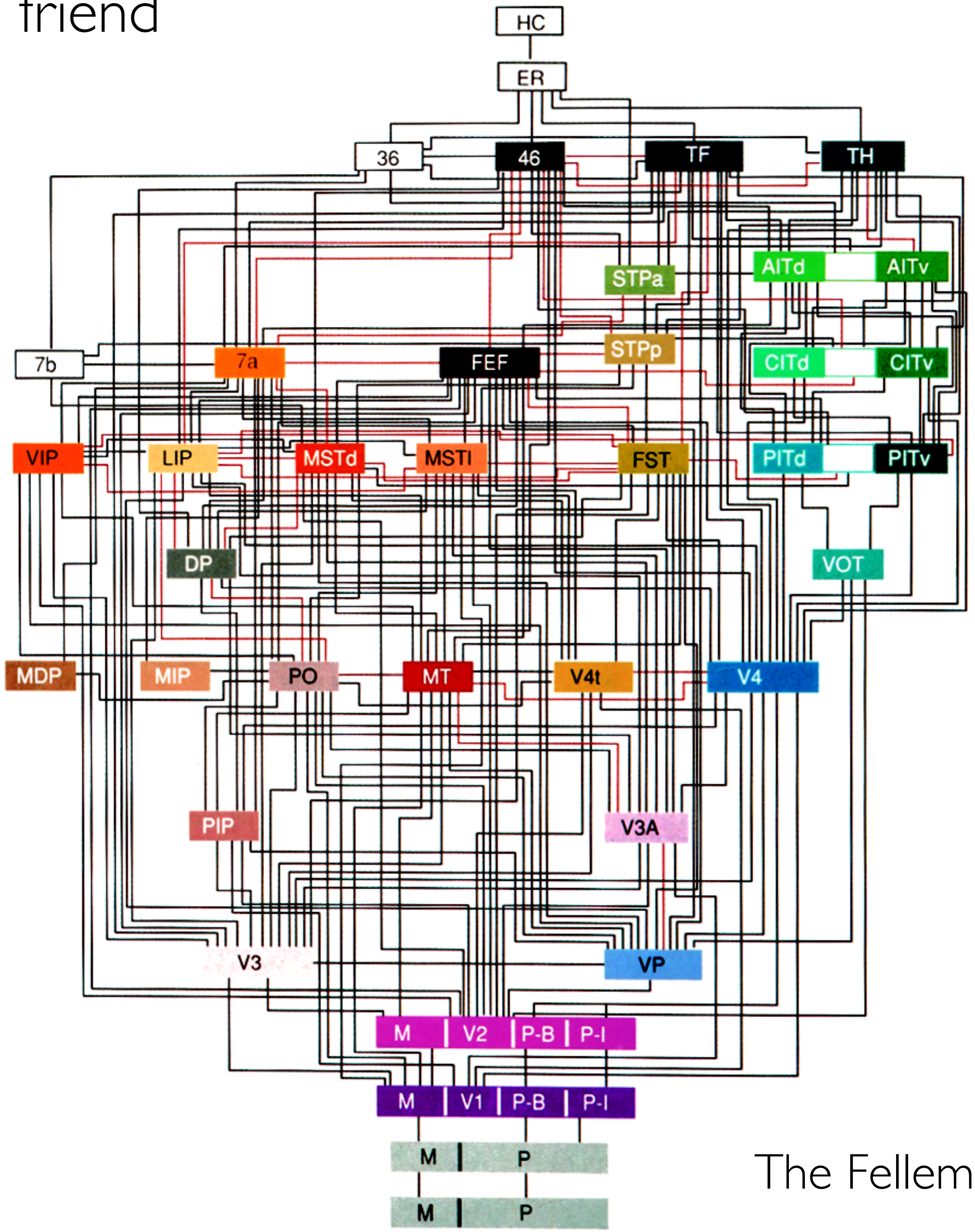
SOLUTION

RECURRENCE and FEEDBACK

SELF-SUPERVISION WORKS GREAT!

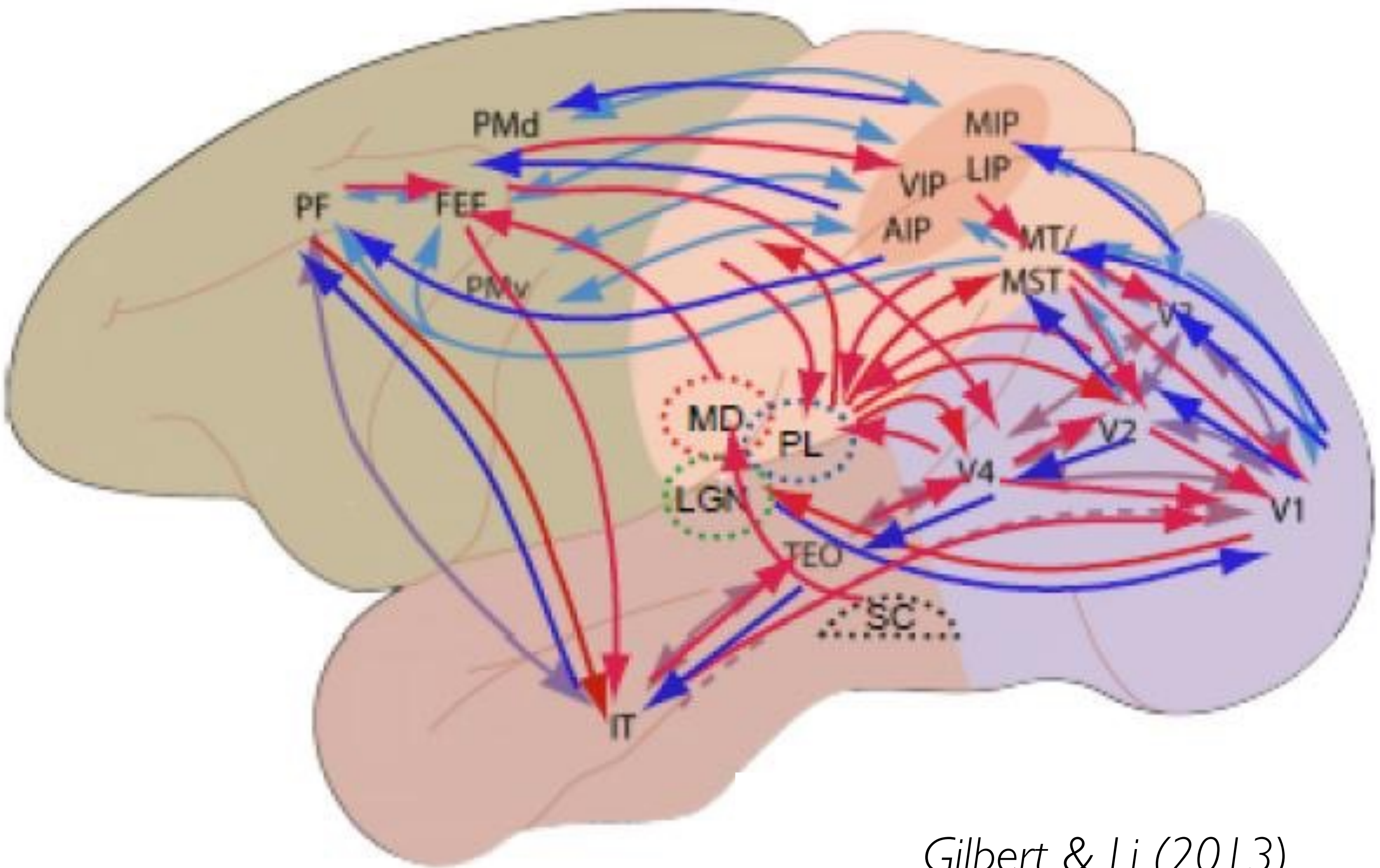
*CAN HANDLE REAL VIDEOSTREAMS TO *SOME* EXTENT*

Our old friend



The Felleman-vanEssen Diagram

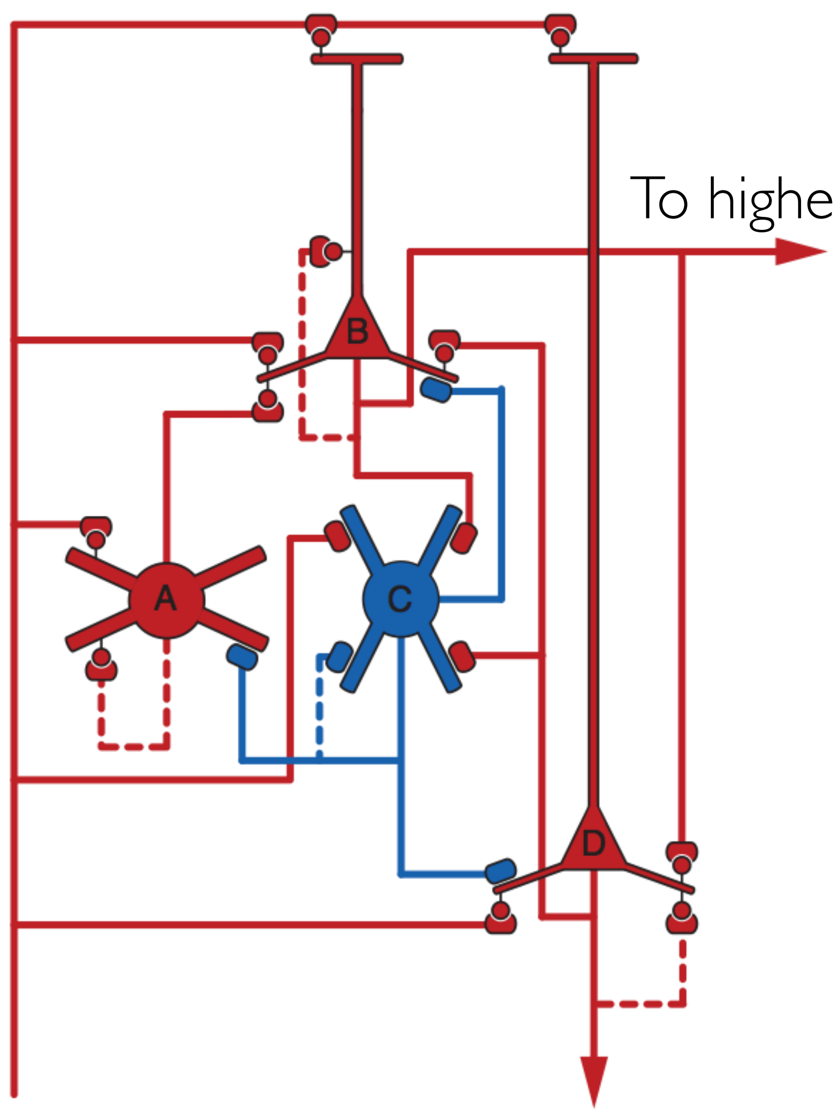
Feedbacks everywhere



Gilbert & Li (2013)

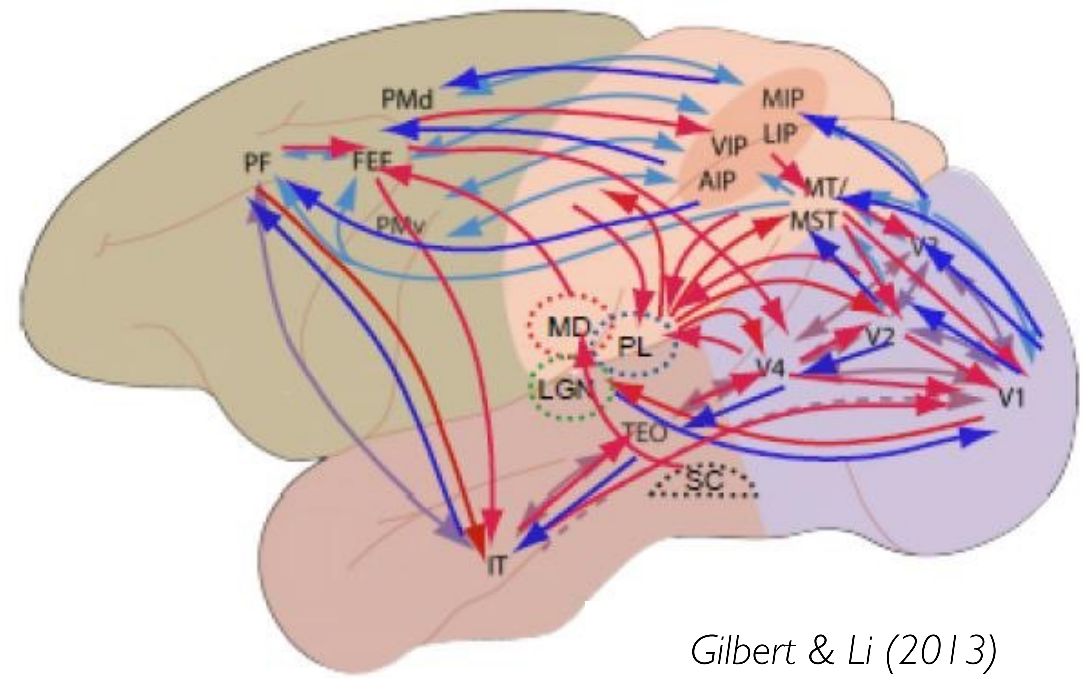
“Real” neural networks are full of feedback

Local recurrence



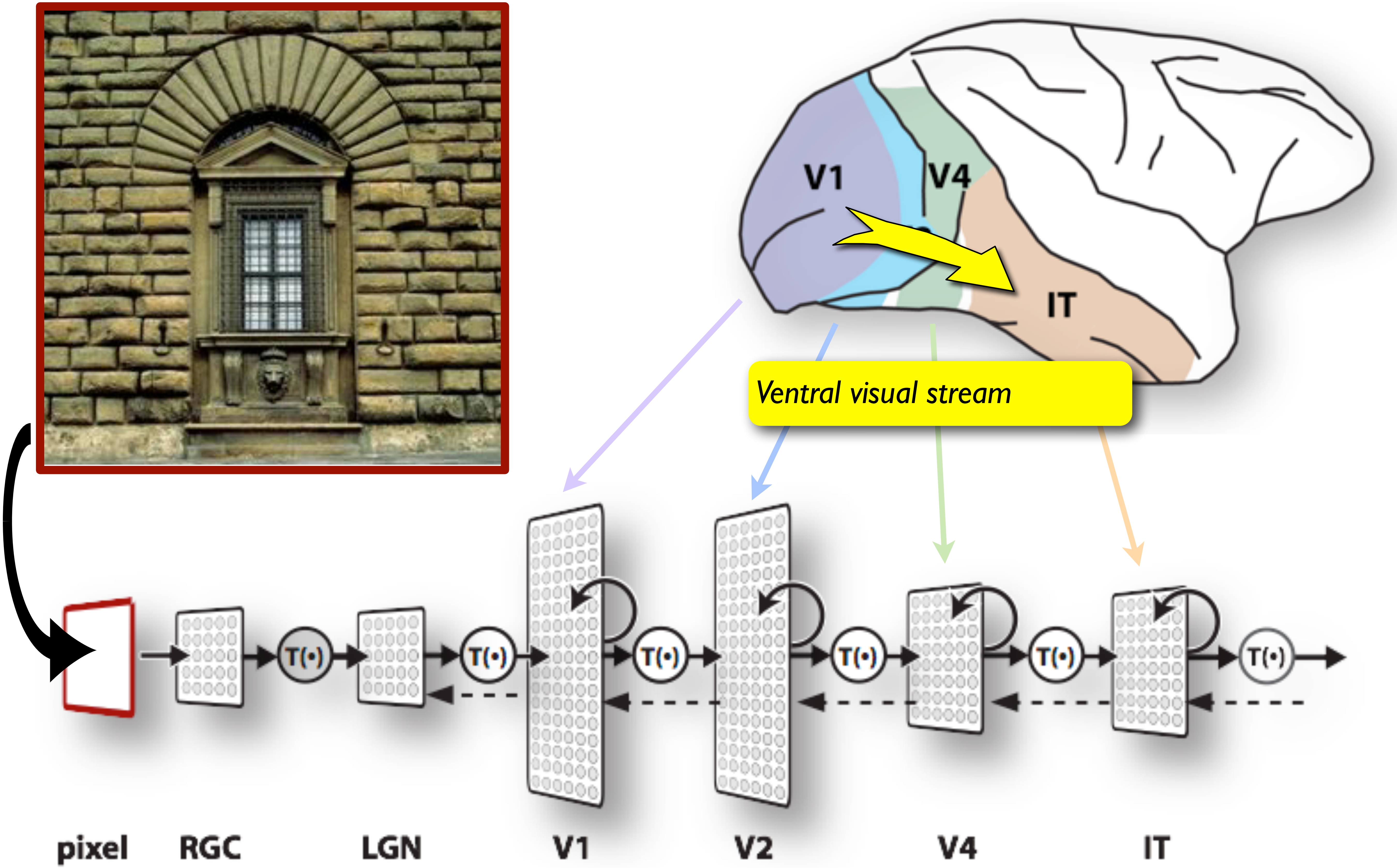
Douglas & Martin (2010)

Long-range connections

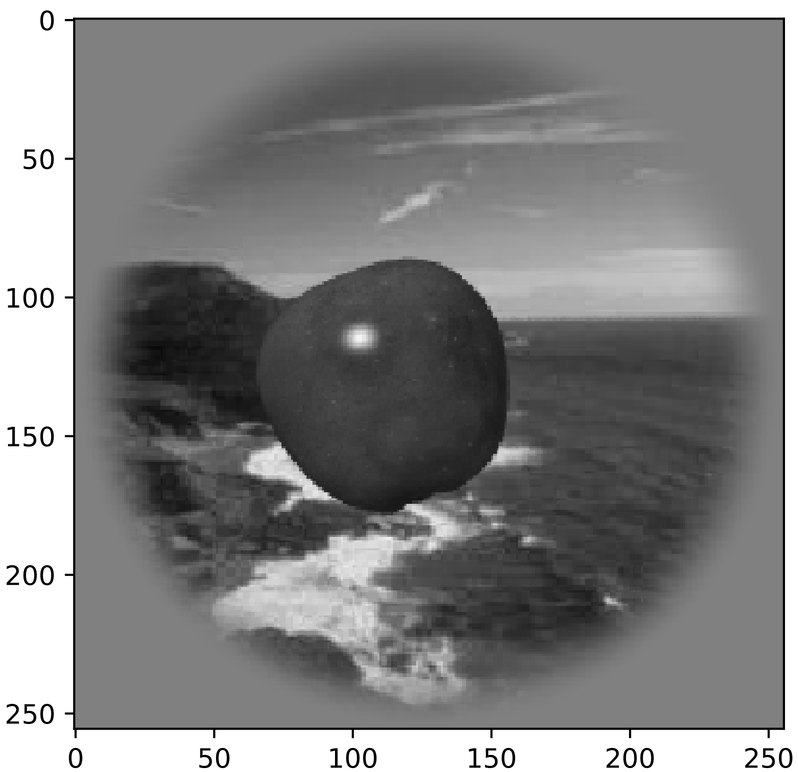


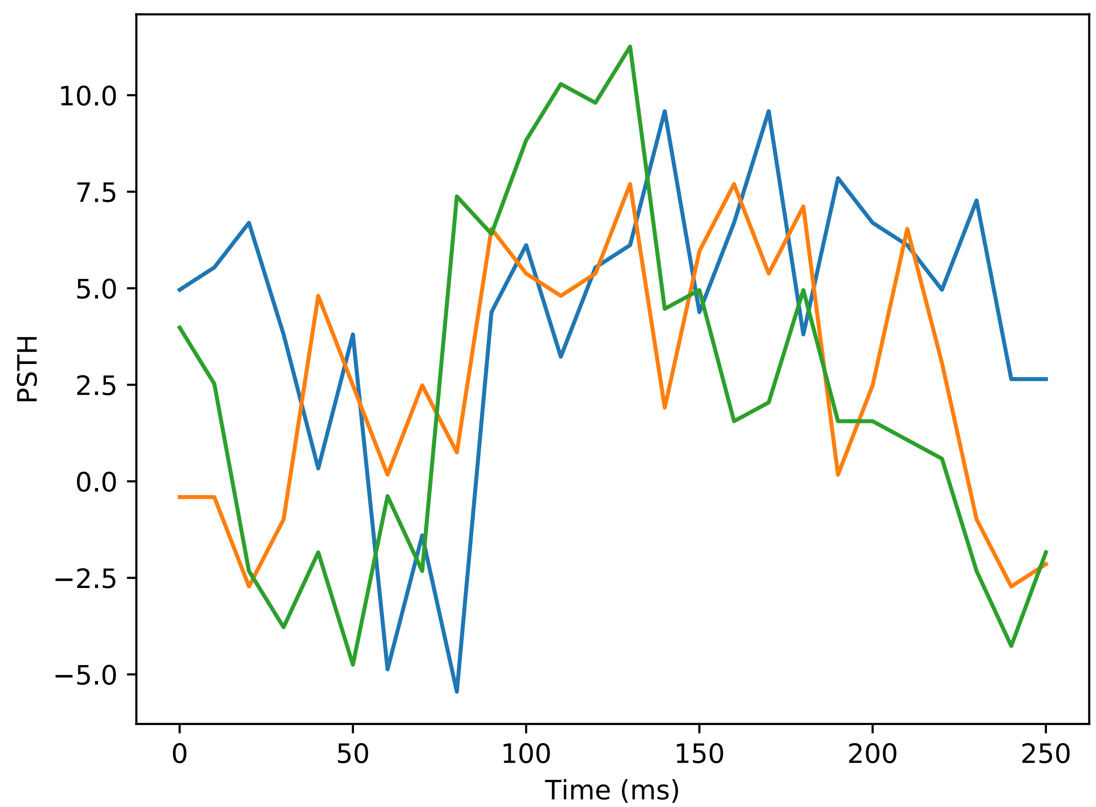
Gilbert & Li (2013)

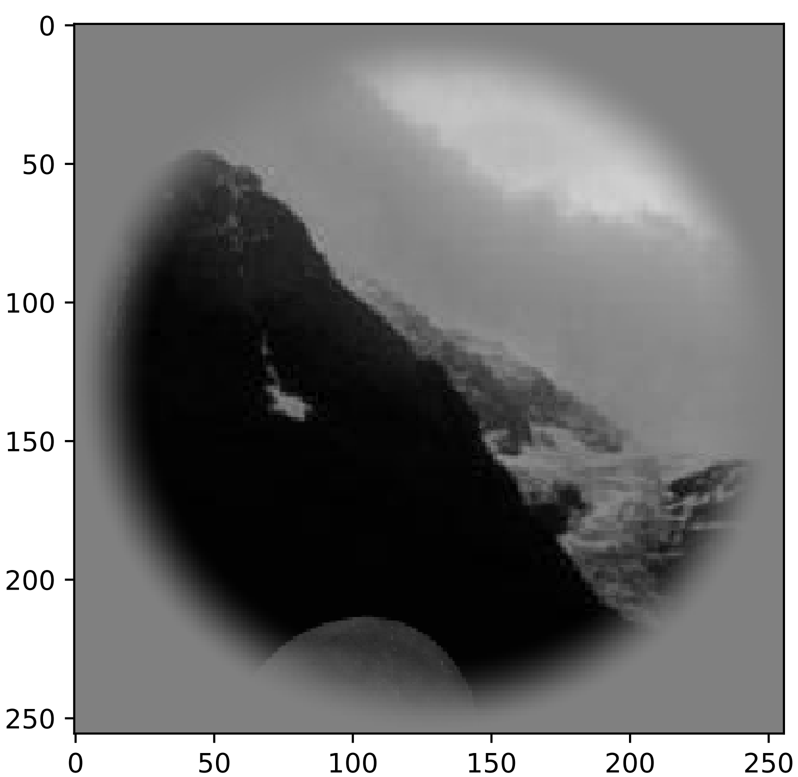
Of course we soft-pedaled them earlier ….

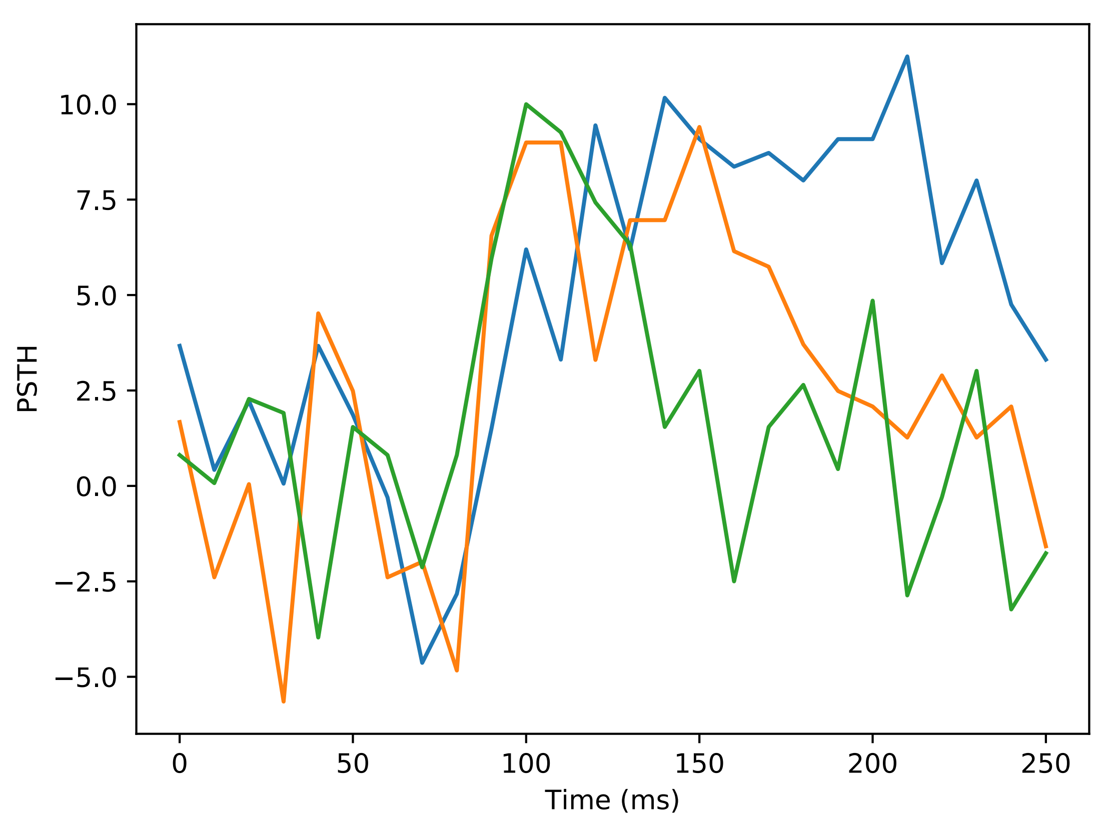


Neural data has dynamics



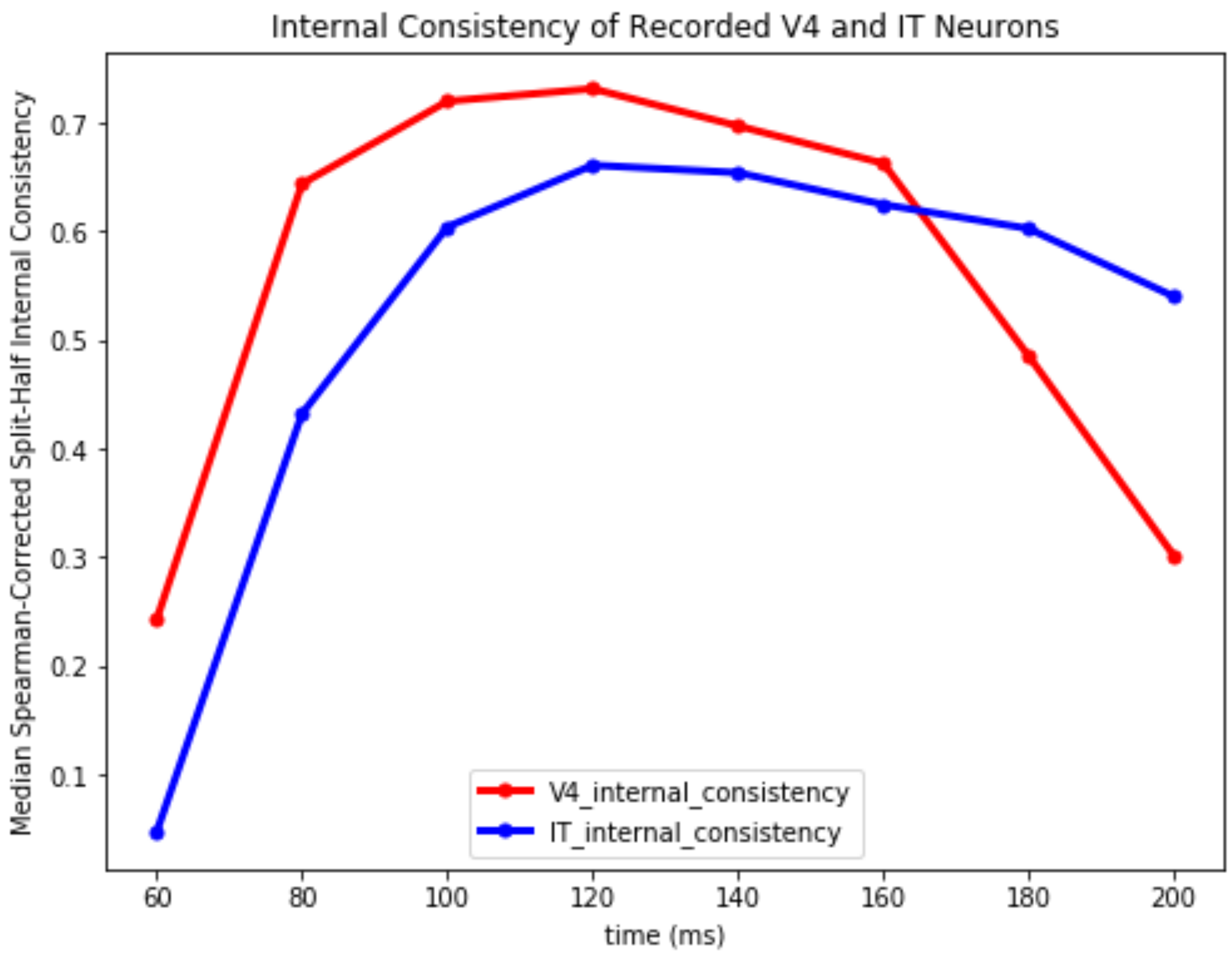




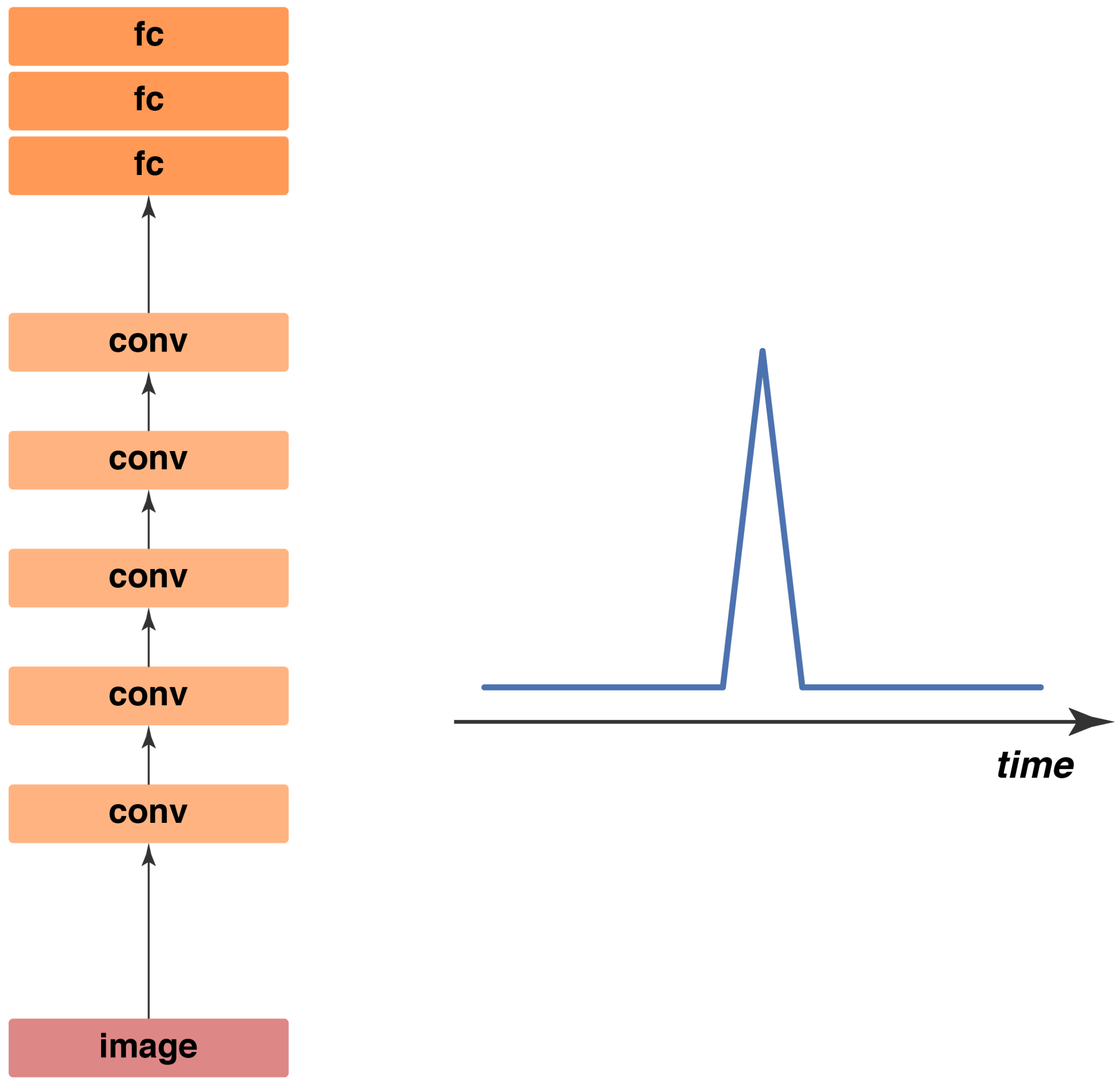


Neural data has dynamics

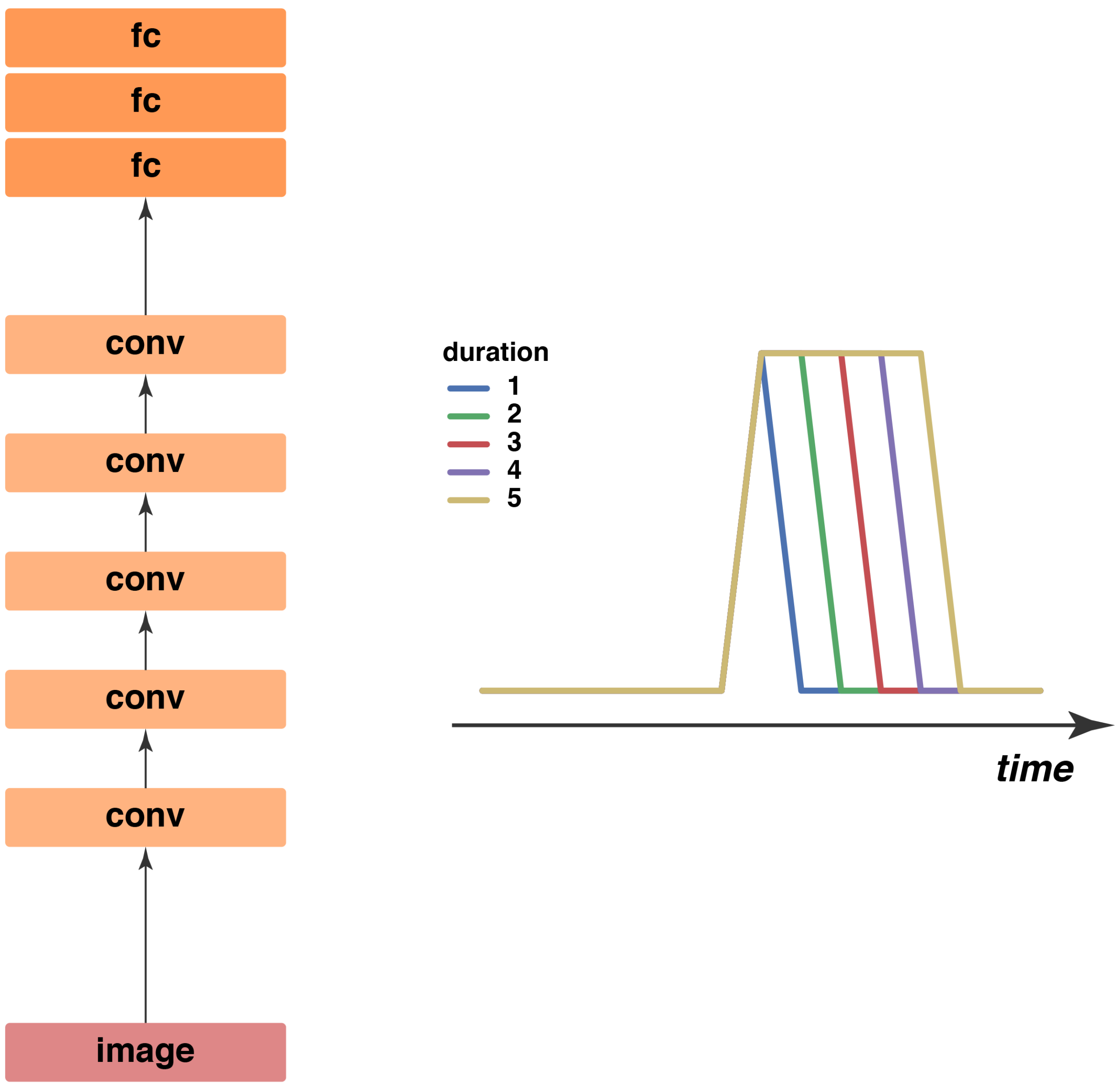
Hierarchical structure can be seen in the dynamics



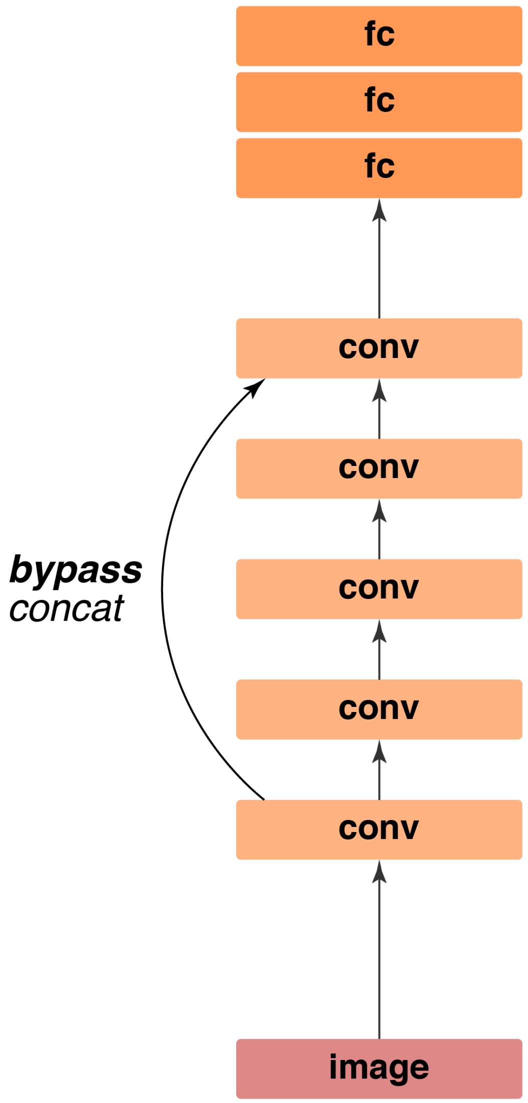
Limitations of Feedforward Structures



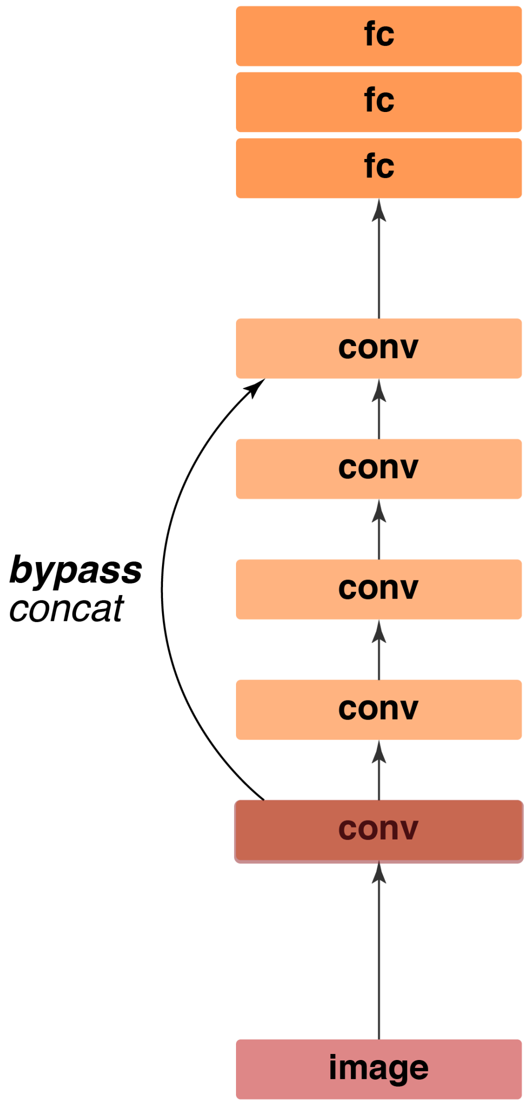
Limitations of Feedforward Structures



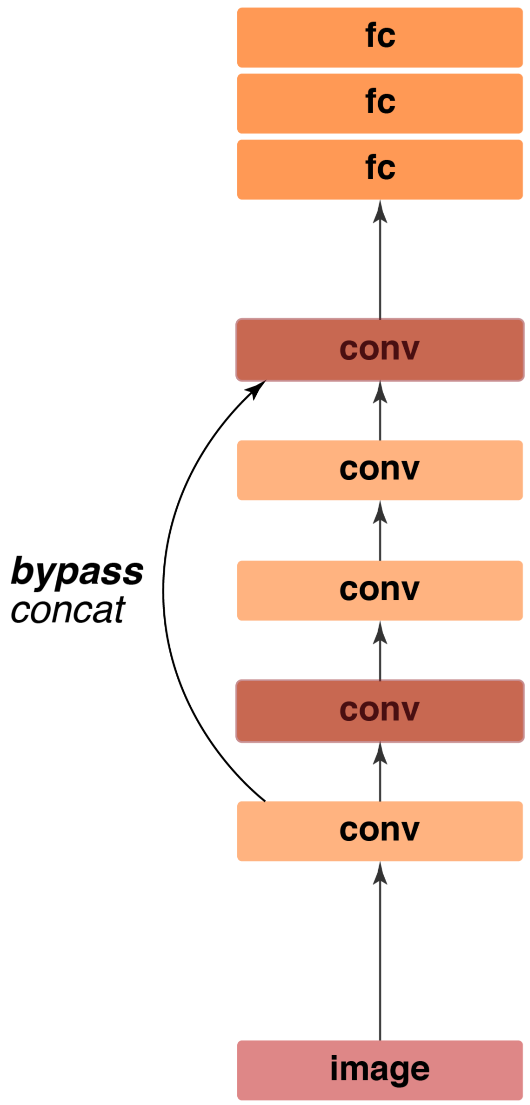
Limitations of Feedforward Structures



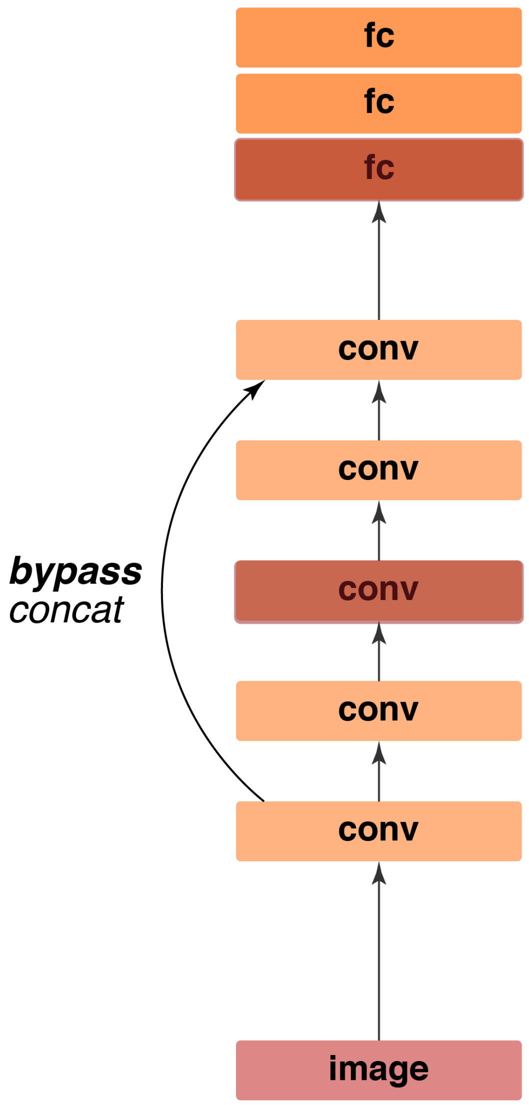
Limitations of Feedforward Structures



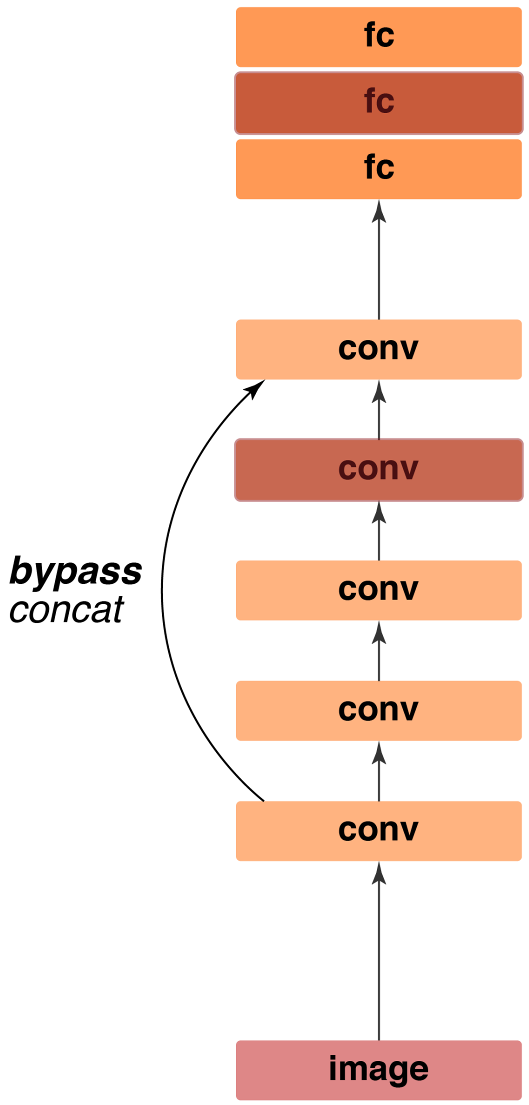
Limitations of Feedforward Structures



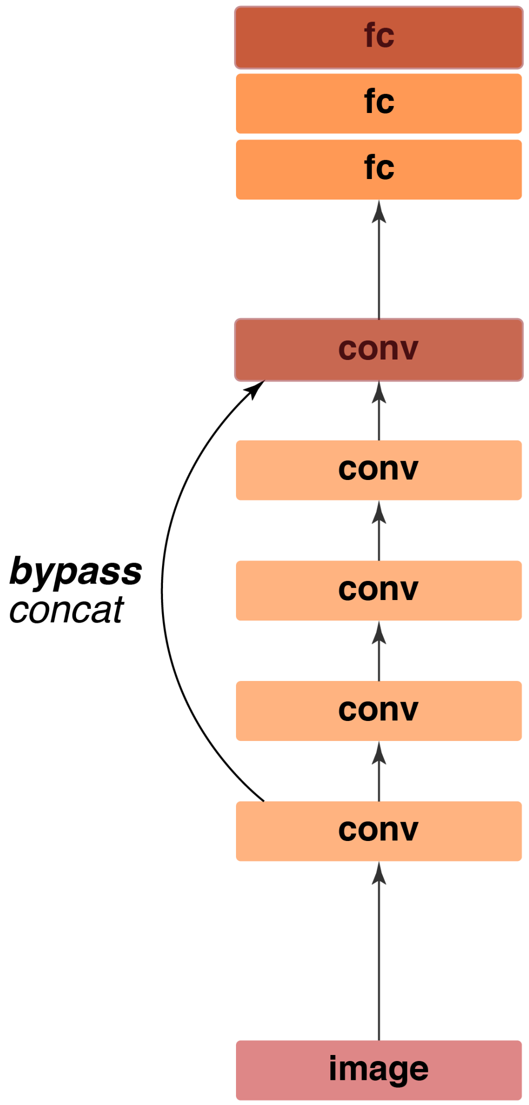
Limitations of Feedforward Structures



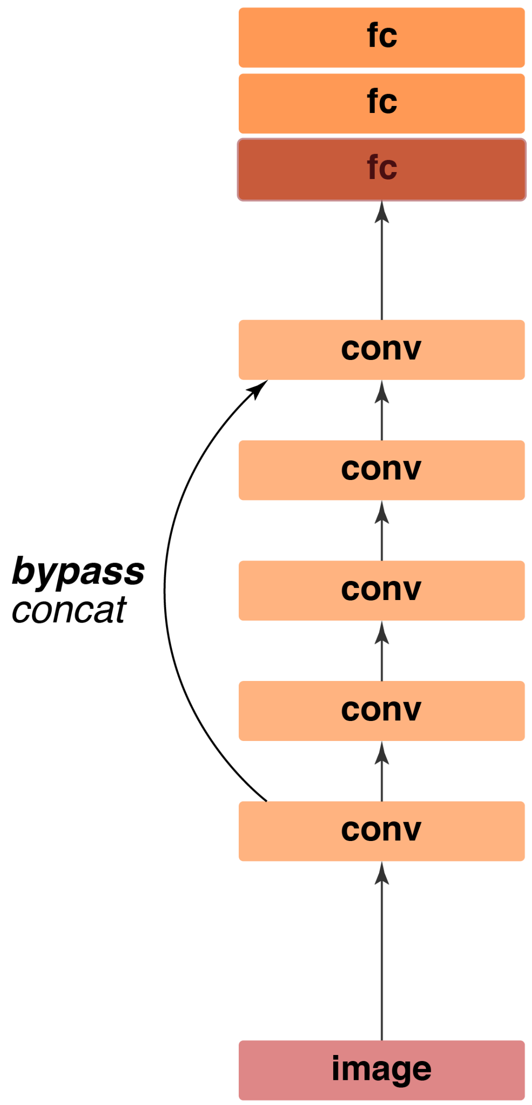
Limitations of Feedforward Structures



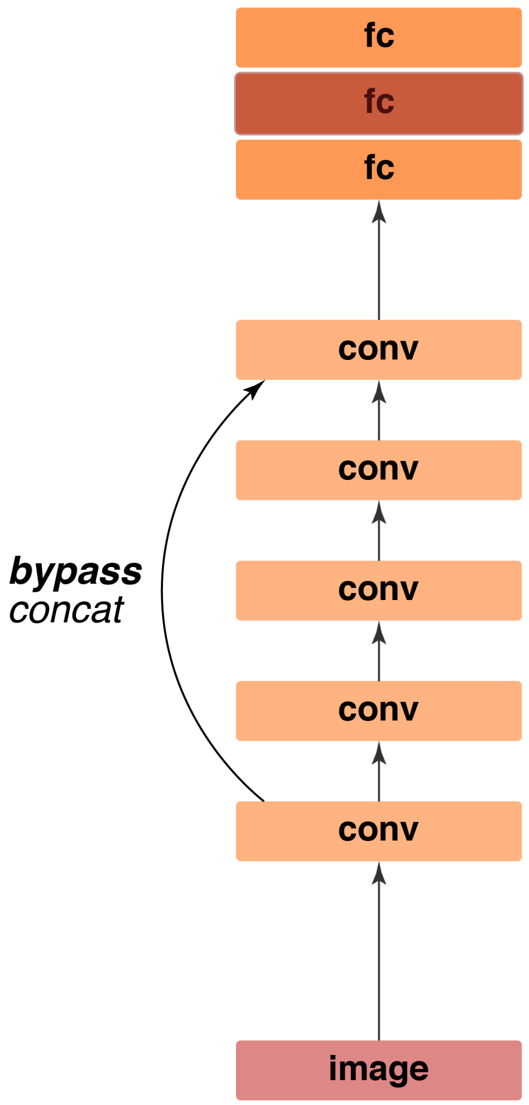
Limitations of Feedforward Structures



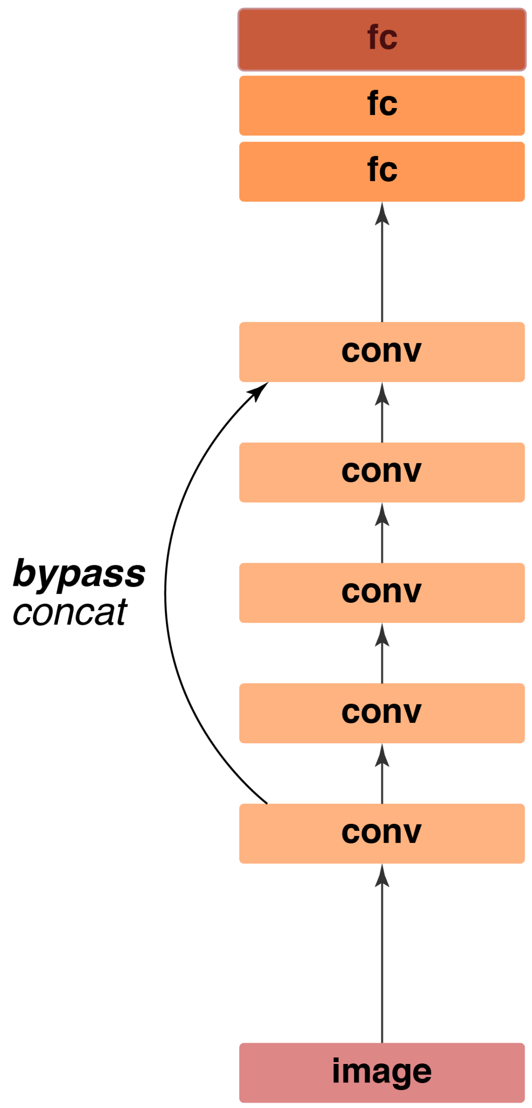
Limitations of Feedforward Structures



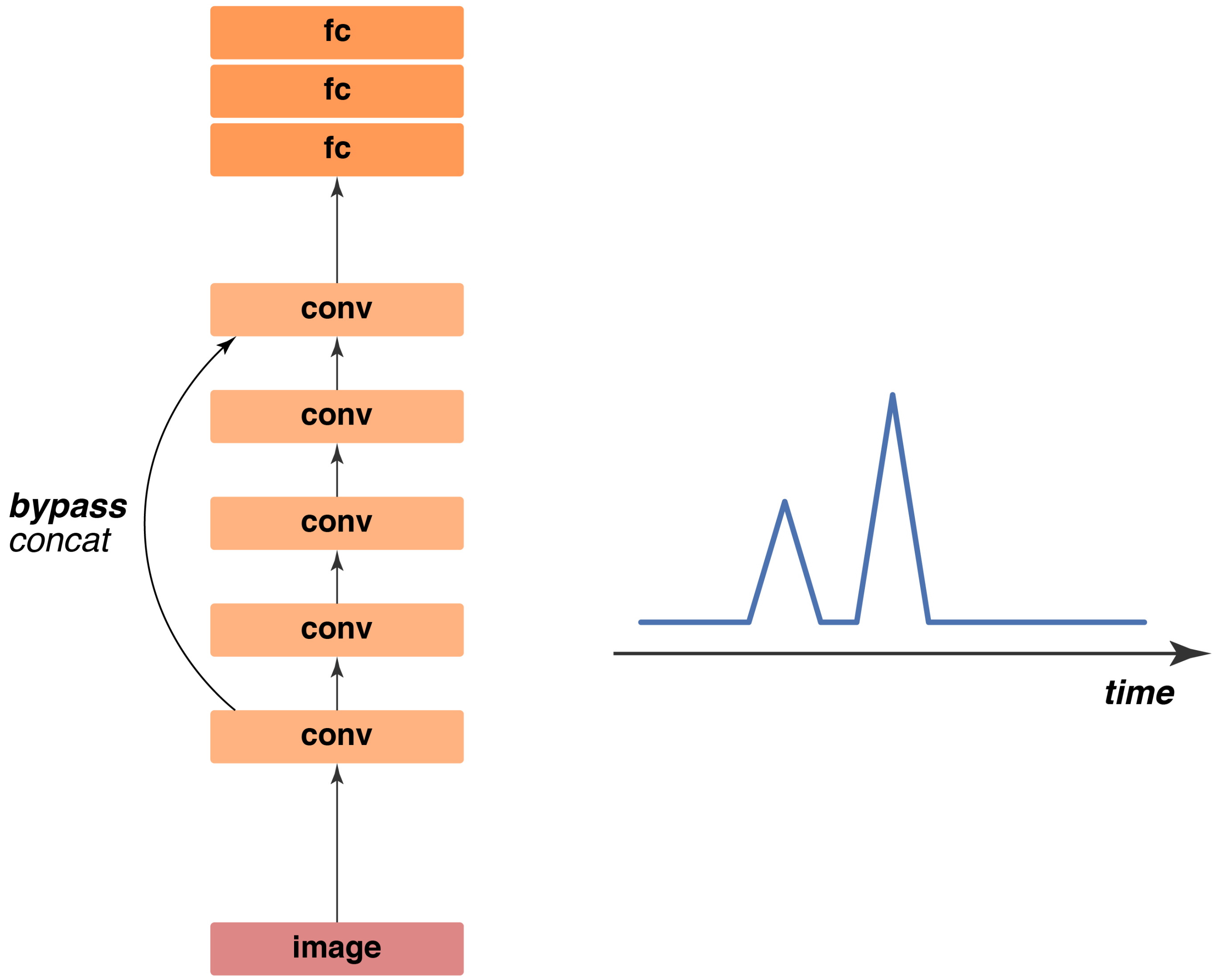
Limitations of Feedforward Structures



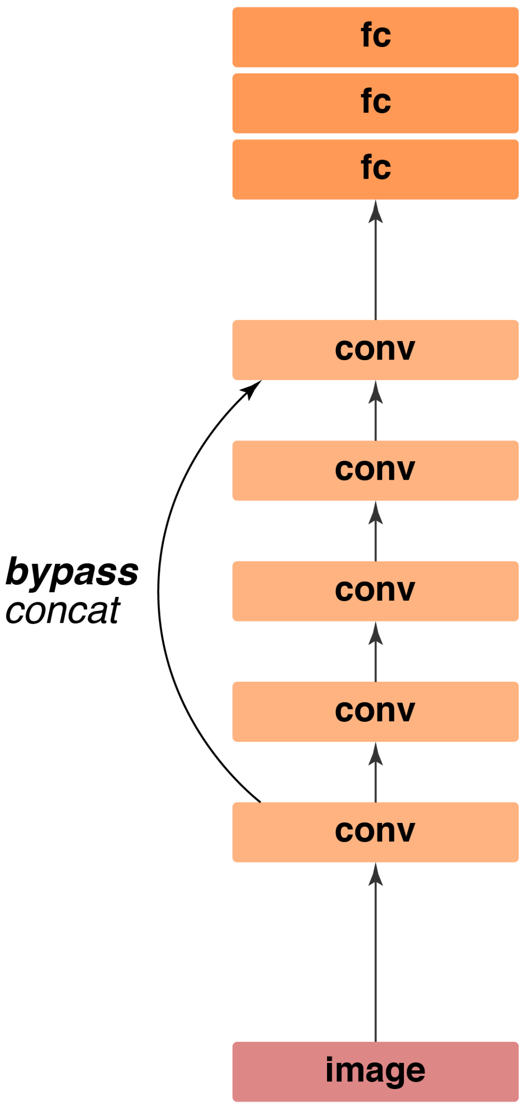
Limitations of Feedforward Structures

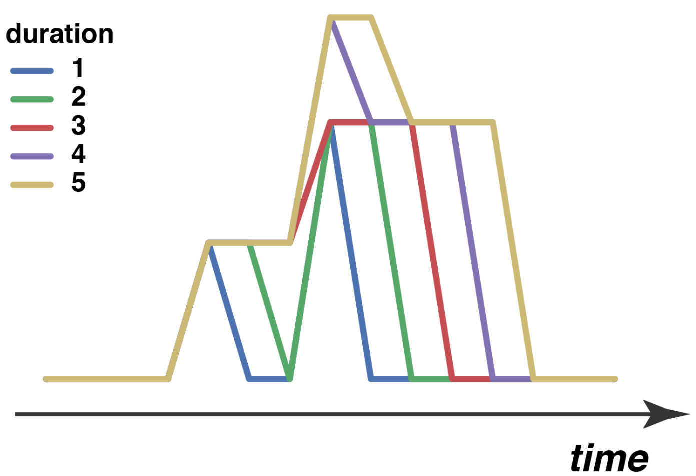


Limitations of Feedforward Structures

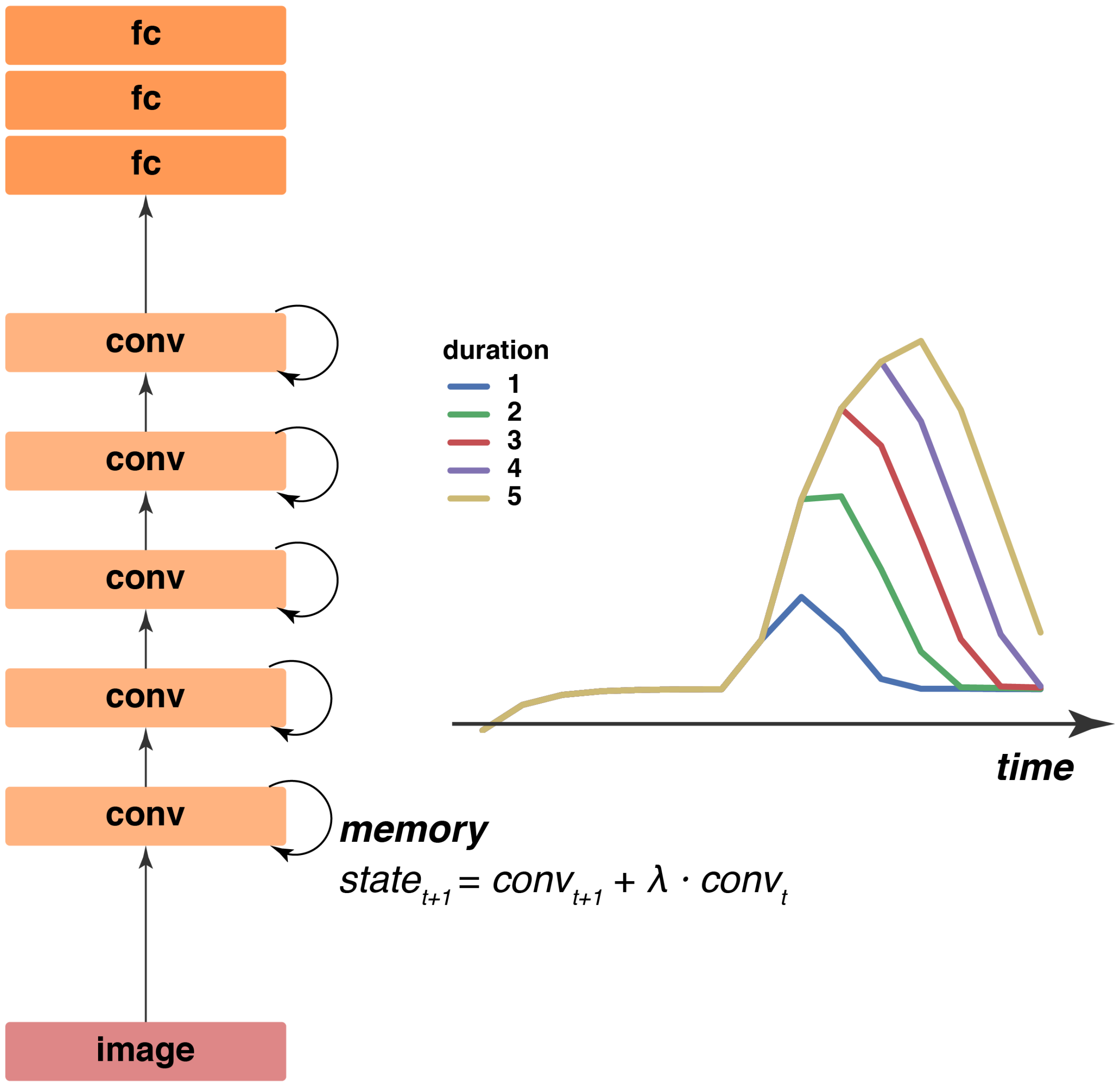


Limitations of Feedforward Structures

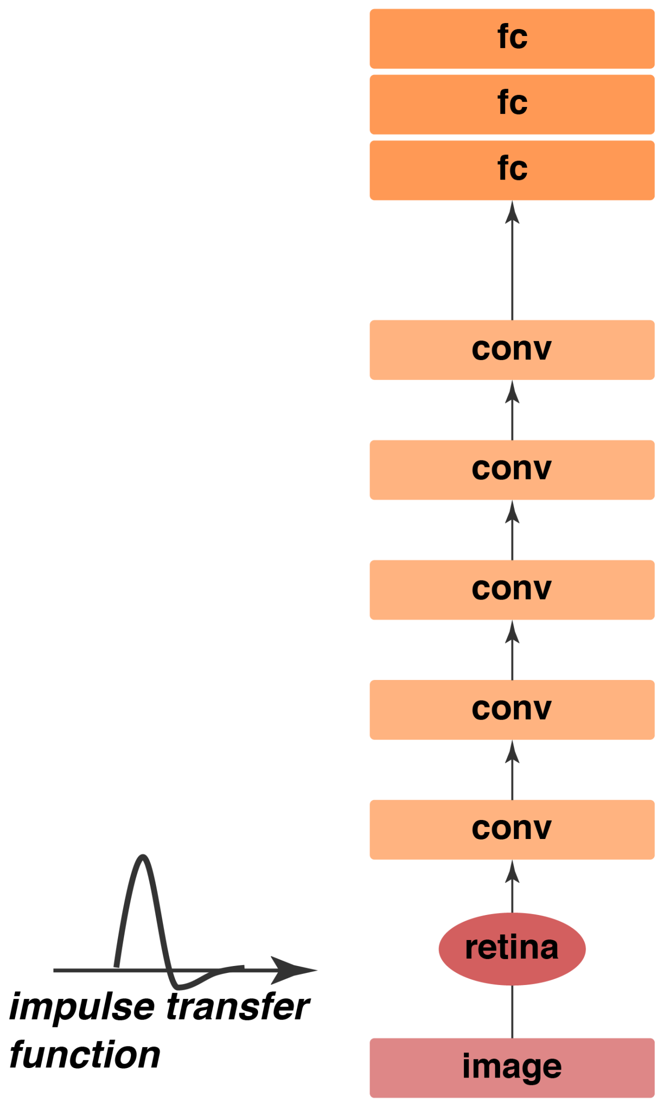




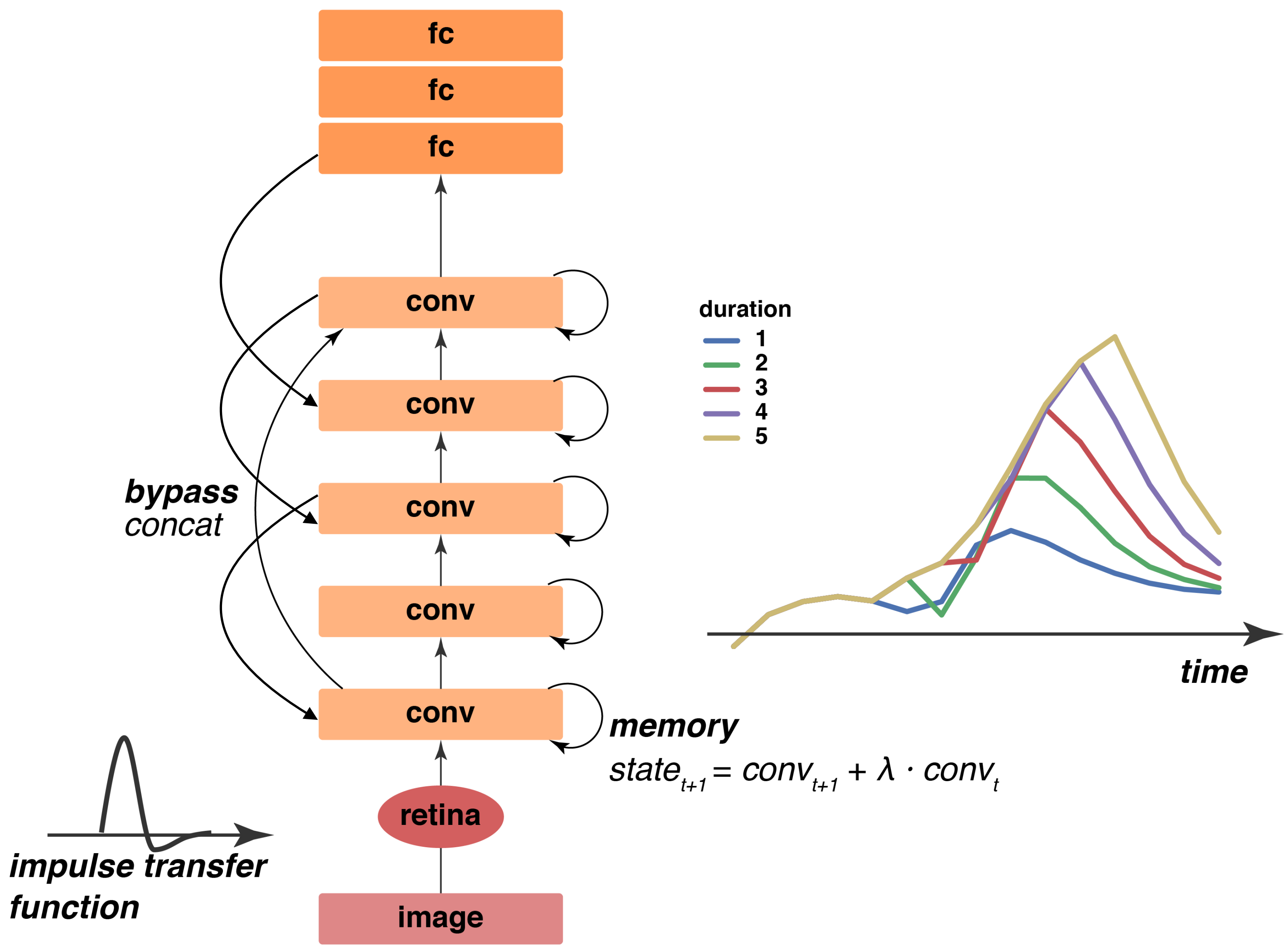
Limitations of Feedforward Structures



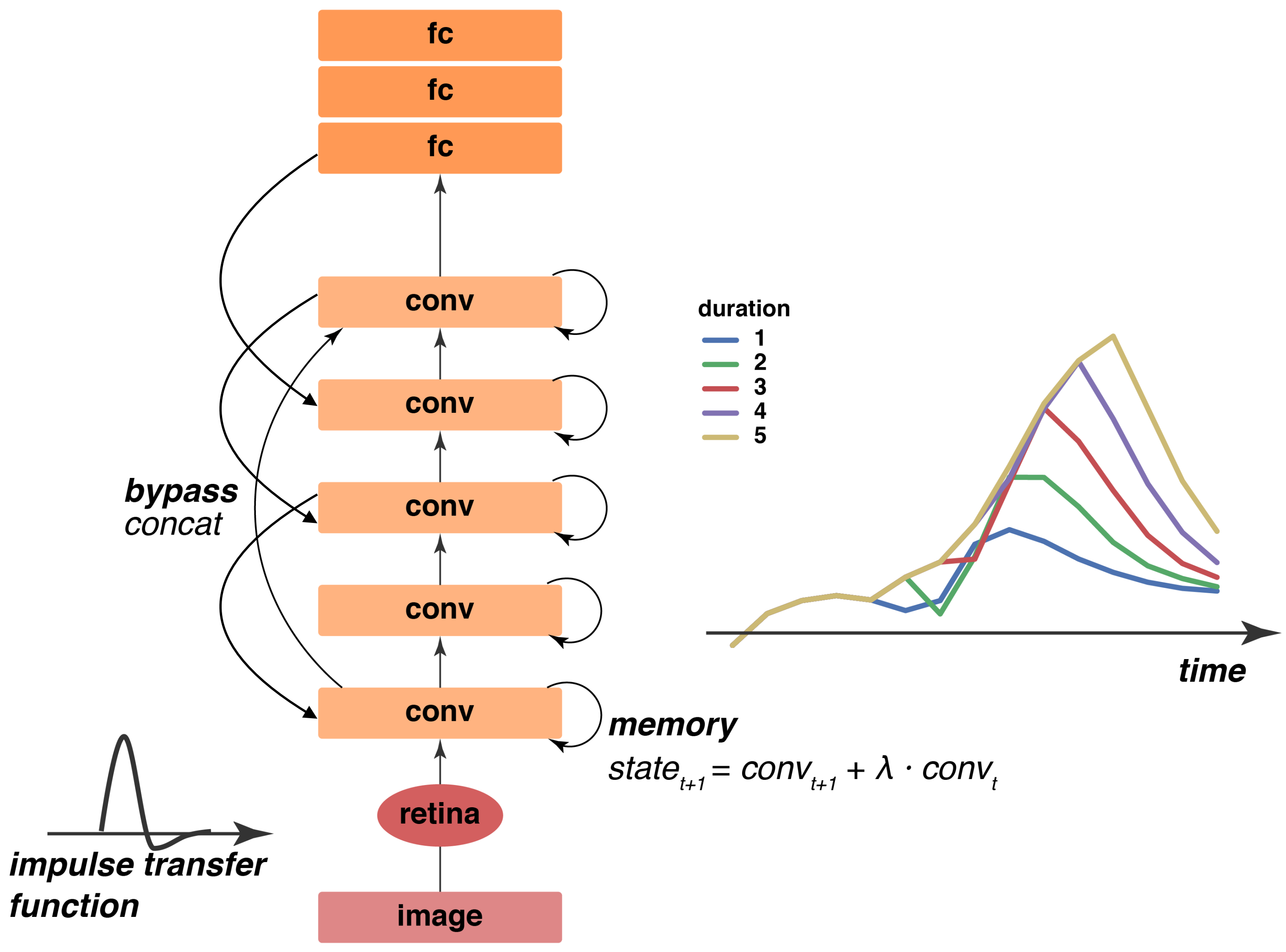
Limitations of Feedforward Structures



Limitations of Feedforward Structures

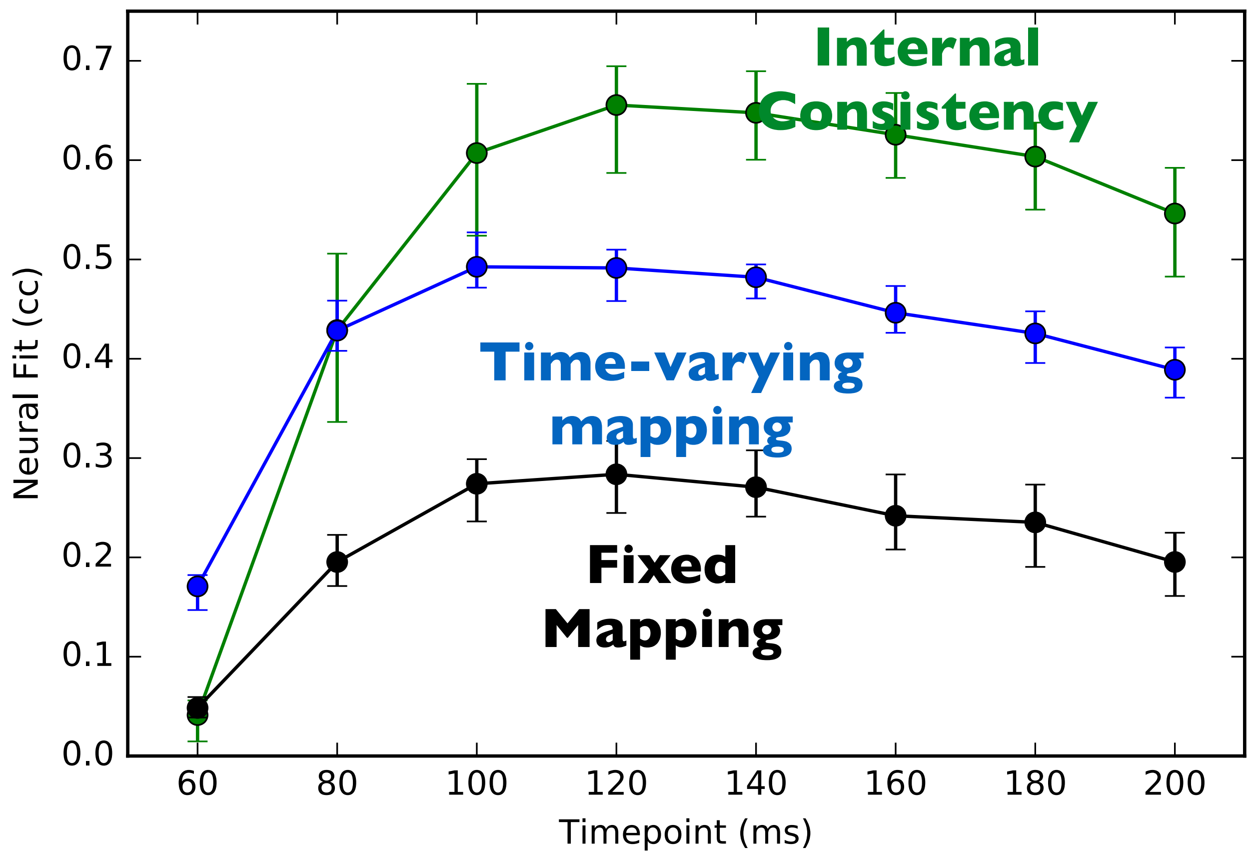


Limitations of Feedforward Structures



Neural data has dynamics

IT trajectories fit with feedforward models



Biological views on function

Top-down influences on visual processing

Charles D. Gilbert[1] and **Wu Li**[2]
[1]The Rockefeller University, 1230 York Avenue, New York, NY 10065
[2]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China

"Vision is an active process, where higher order cognitive influences affect the operations performed by cortical neurons."

Biological views on function

Top-down influences on visual processing

Charles D. Gilbert[1] and **Wu Li**[2]
[1]The Rockefeller University, 1230 York Avenue, New York, NY 10065
[2]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China

“Vision is an active process, where higher order cognitive influences affect the operations performed by cortical neurons.”

“Top-down influences include various forms of attention, including spatial, object oriented and feature oriented attention.”

Biological views on function

Top-down influences on visual processing

Charles D. Gilbert[1] and **Wu Li**[2]
[1]The Rockefeller University, 1230 York Avenue, New York, NY 10065
[2]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China

"Vision is an active process, where higher order cognitive influences affect the operations performed by cortical neurons."

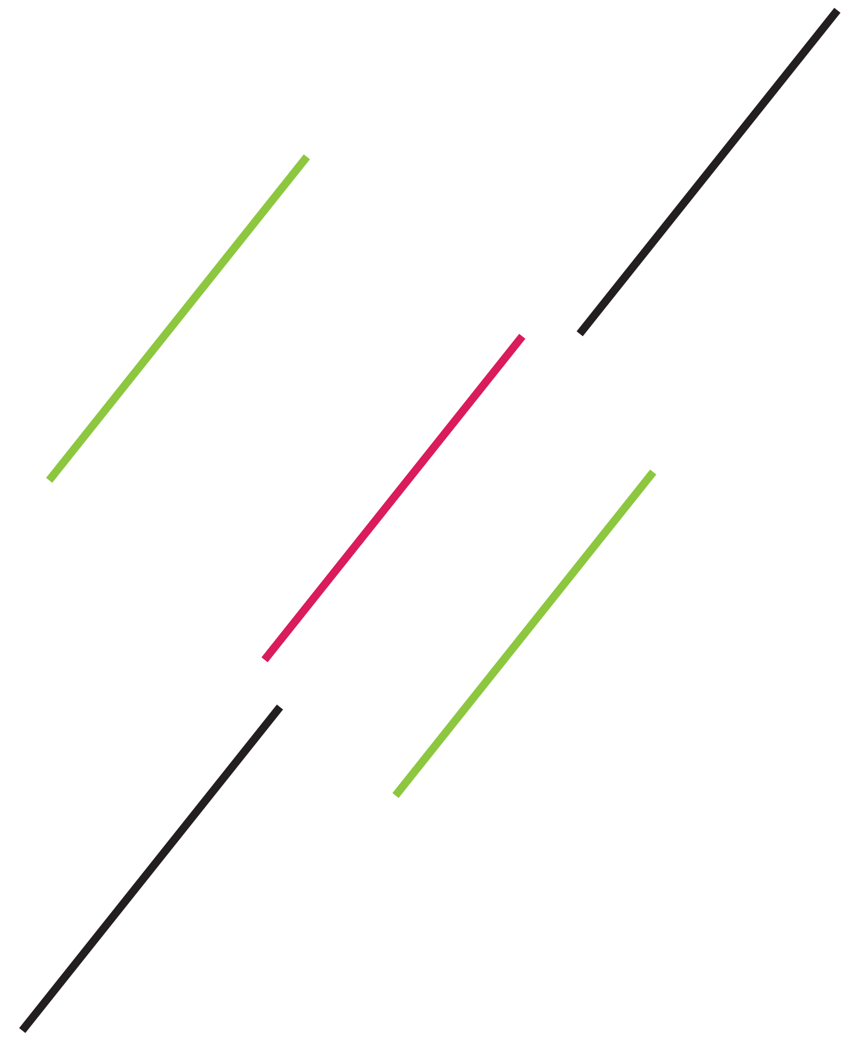
"Top-down influences include various forms of attention, including spatial, object oriented and feature oriented attention."

"Top-down influences …. [also include] perceptual task, object expectation, scene segmentation, efference copy, working memory, and the encoding and recall of learned information."

Biological views on function

Task-dependent changes in neural tuning and information content in V1

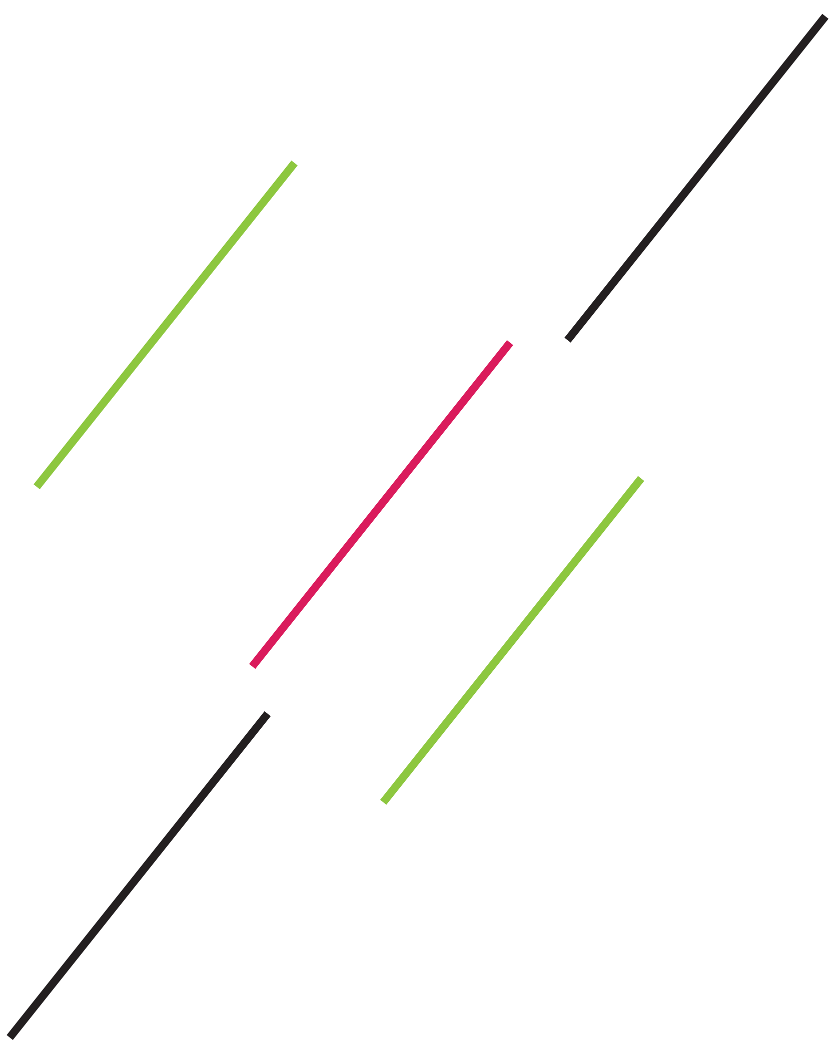
T1: "Which green line is closer to the red?"



Biological views on function

Task-dependent changes in neural tuning and information content in V1

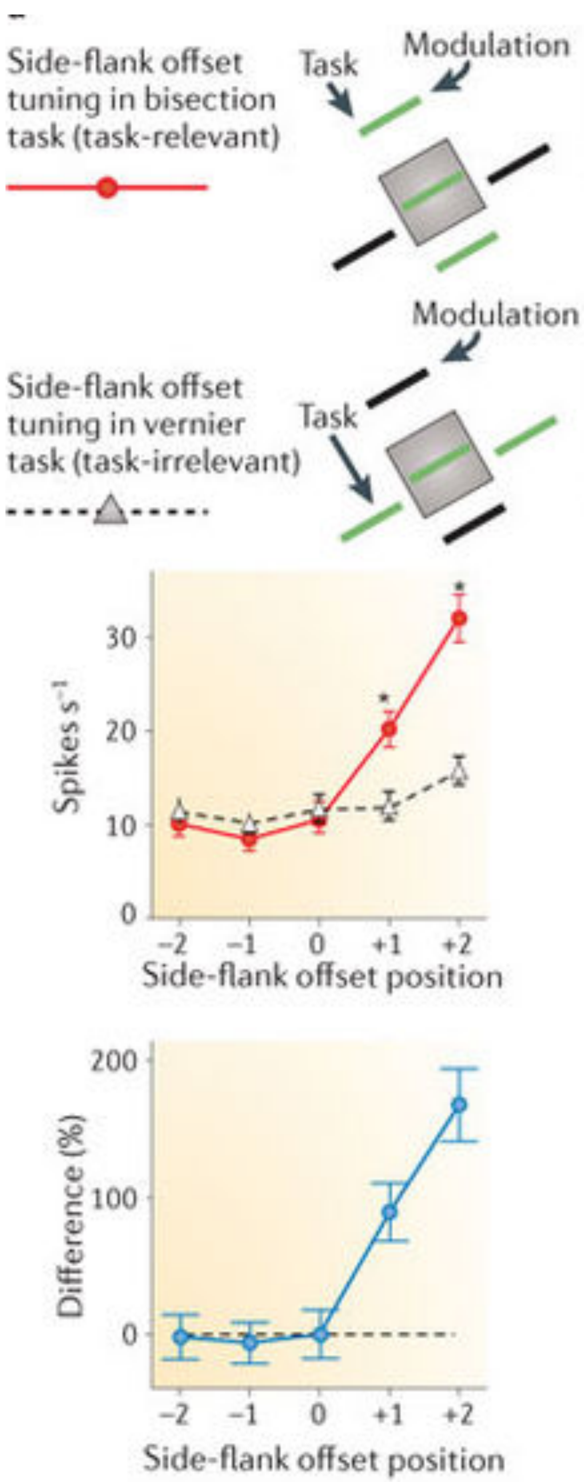
T1: "Which green line is closer to the red?"



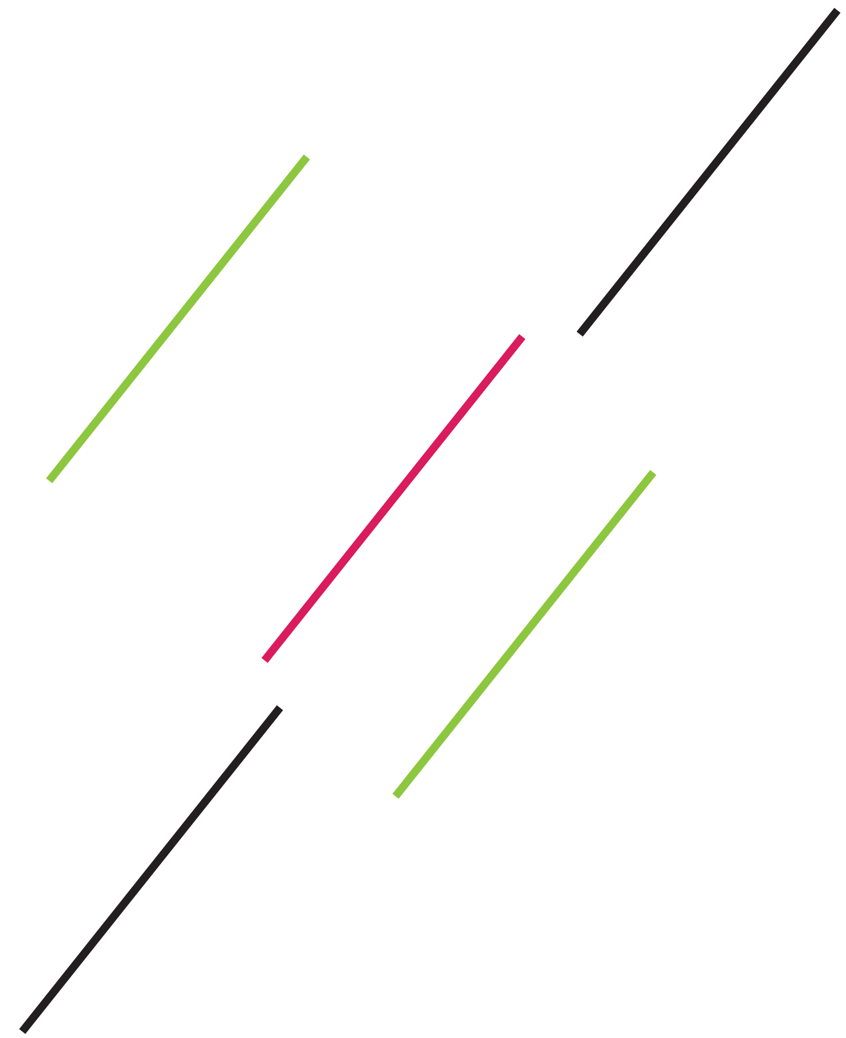
T2: "Which black line is closer to the red?"

Biological views on function

Task-dependent changes in neural tuning and information content in V1



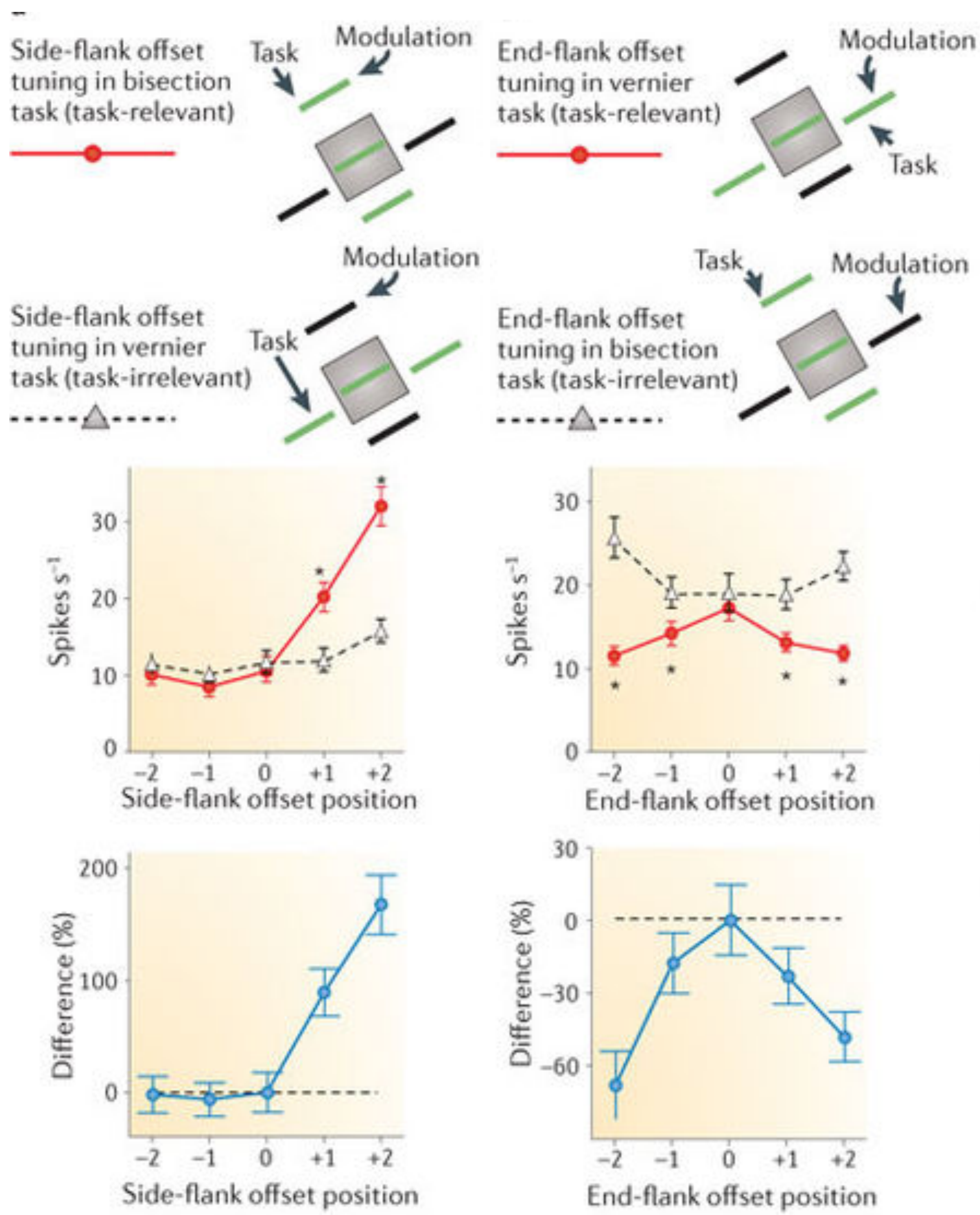
T1: "Which green line is closer to the red?"



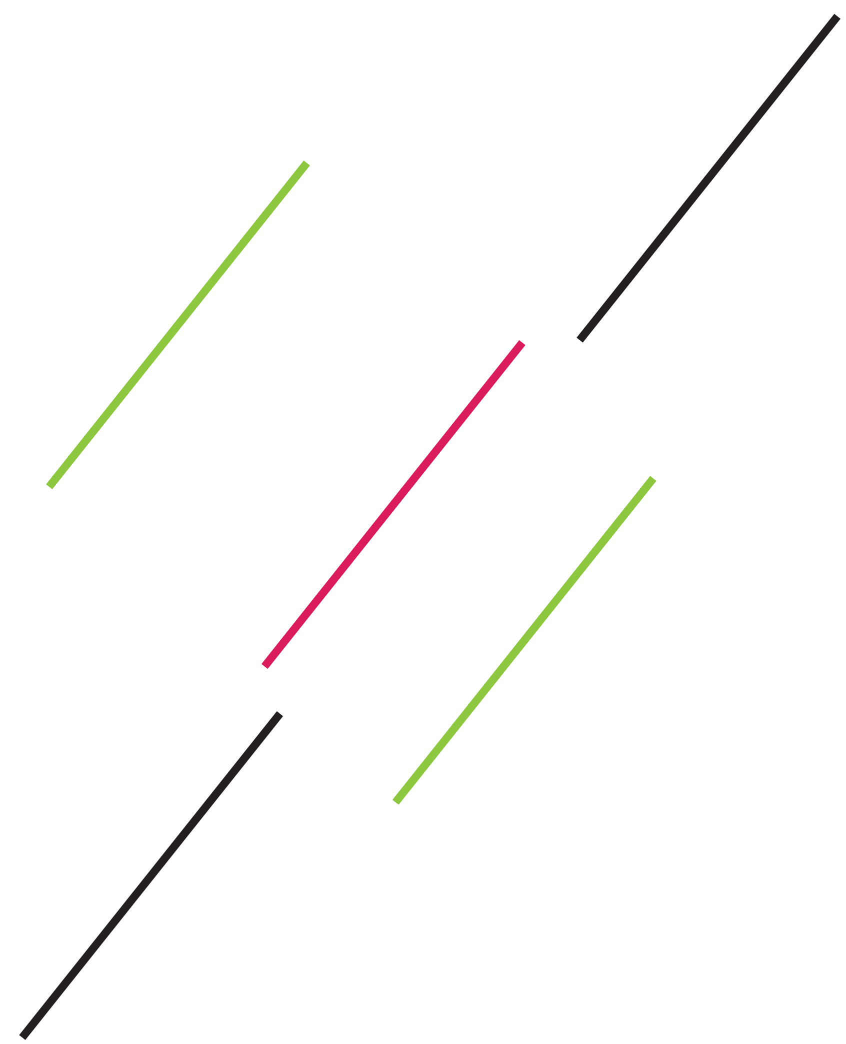
T2: "Which black line is closer to the red?"

Biological views on function

Task-dependent changes in neural tuning and information content in V1



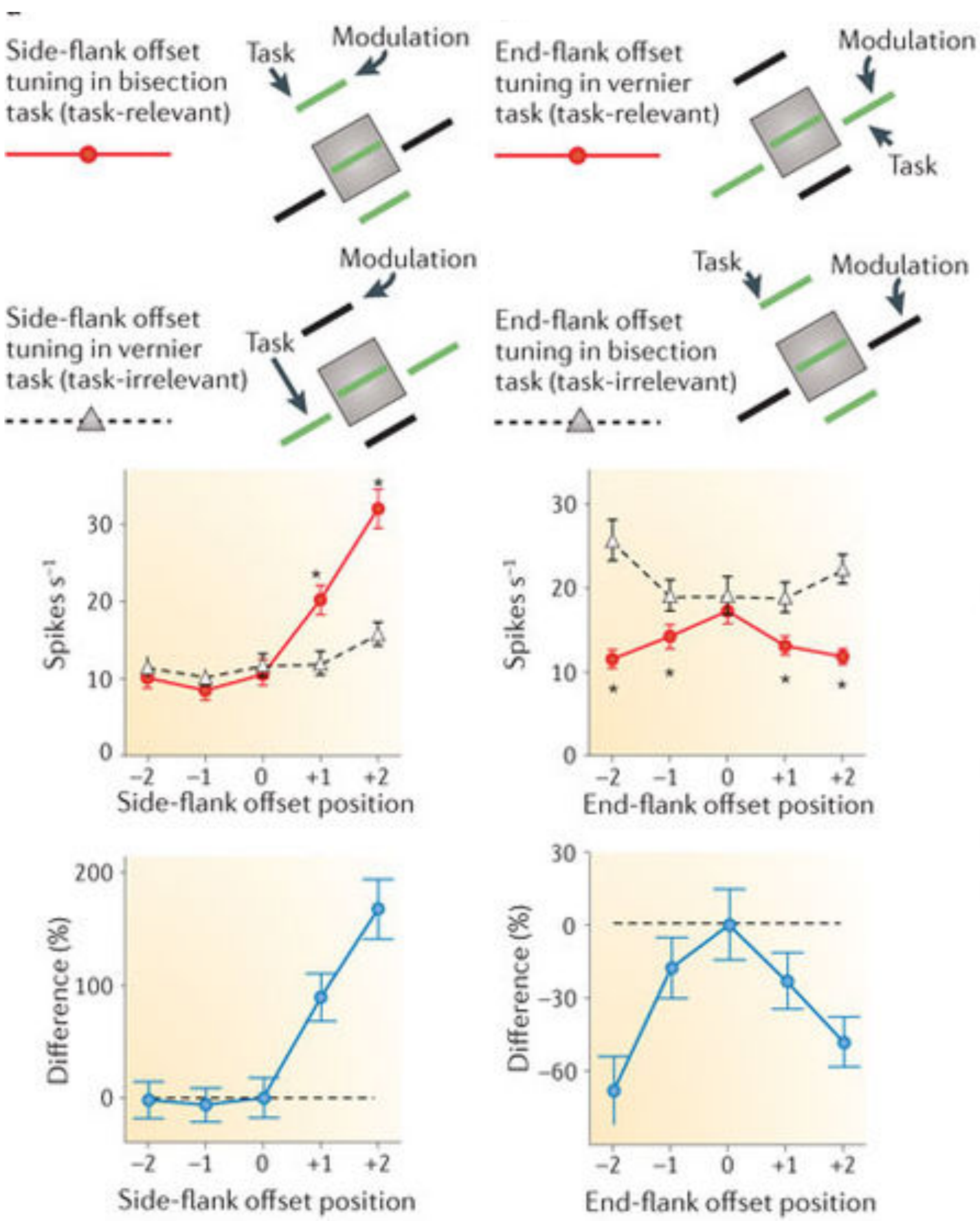
T1: "Which green line is closer to the red?"



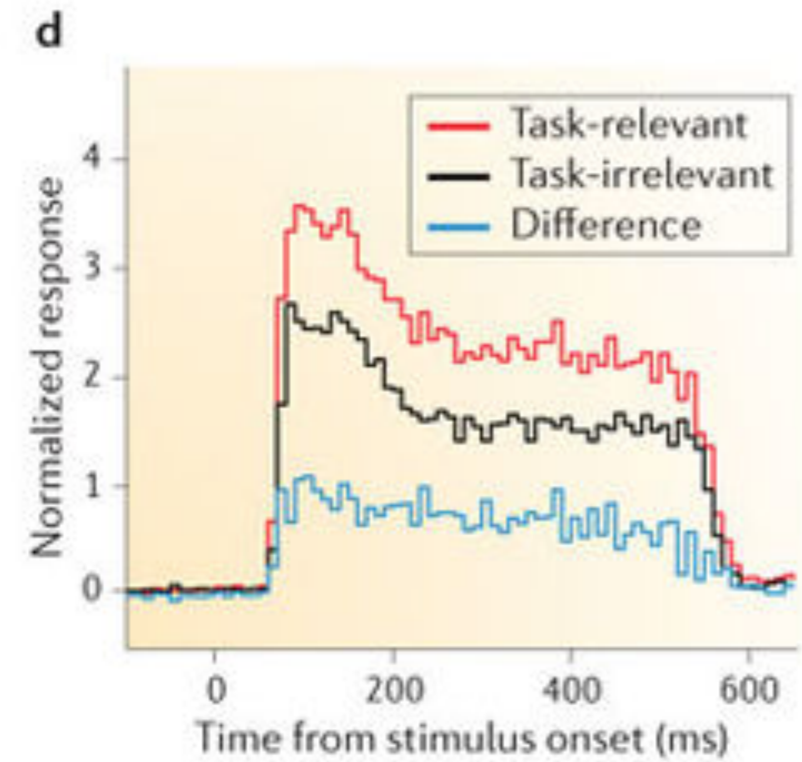
T2: "Which black line is closer to the red?"

Biological views on function

Task-dependent changes in neural tuning and information content in V1



T1: "Which green line is closer to the red?"



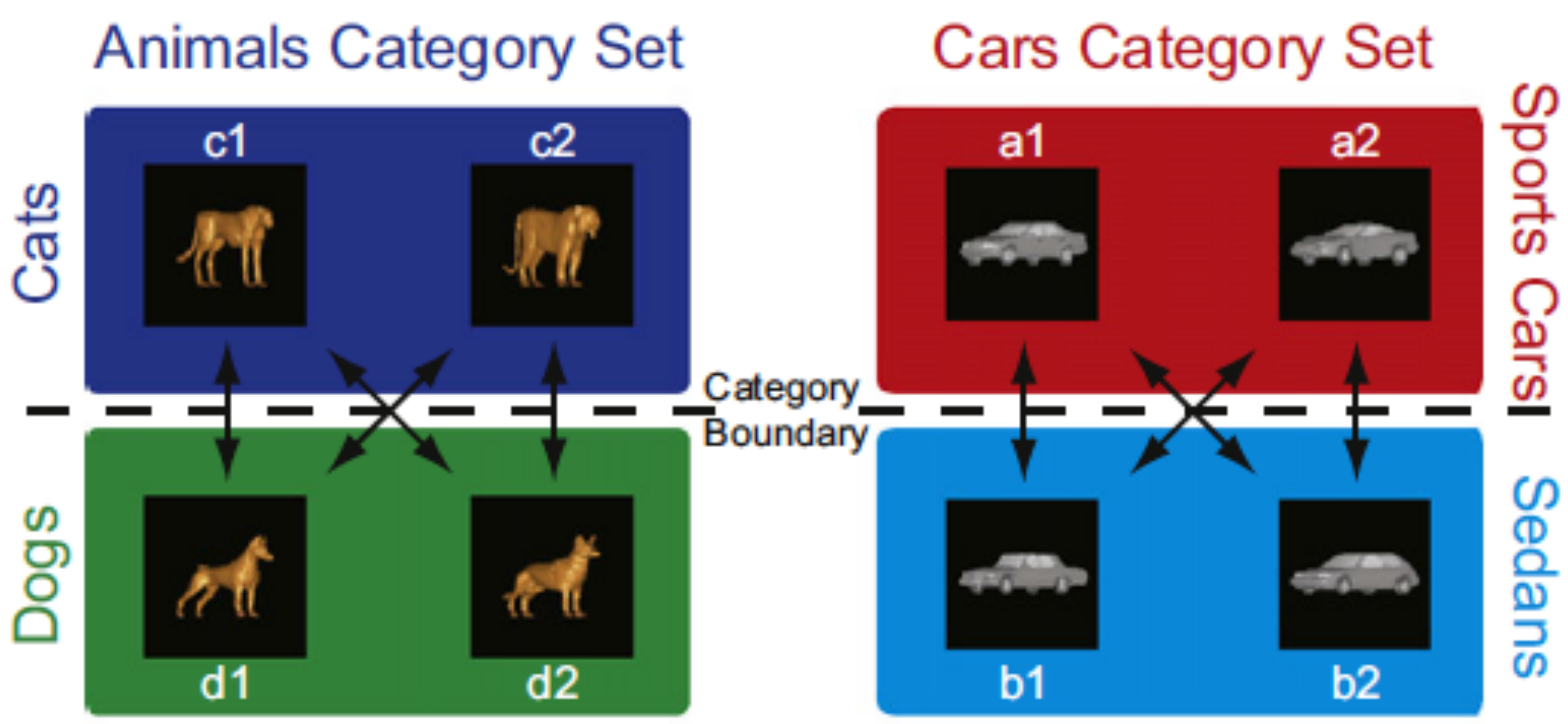
T2: "Which black line is closer to the red?"

Biological views on function

Task-dependent changes in neural tuning and information content in IT

Comparison of Primate Prefrontal and Premotor Cortex Neuronal Activity during Visual Categorization

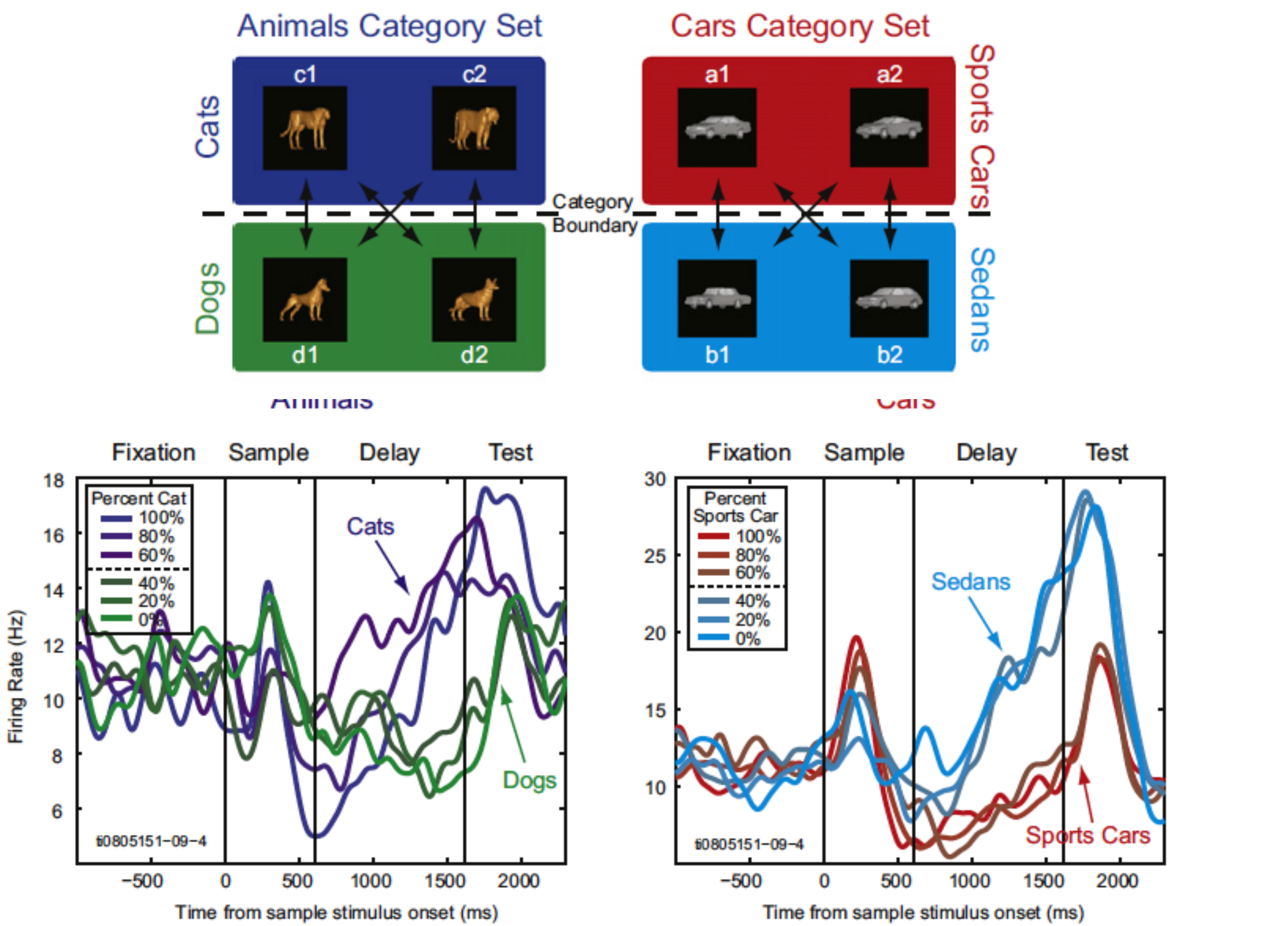
Jason A. Cromer, Jefferson E. Roy, Timothy J. Buschman, and Earl K. Miller



Cromer & Miller 2010

Biological views on function

Task-dependent changes in neural tuning and information content in IT



Cromer & Miller 2010

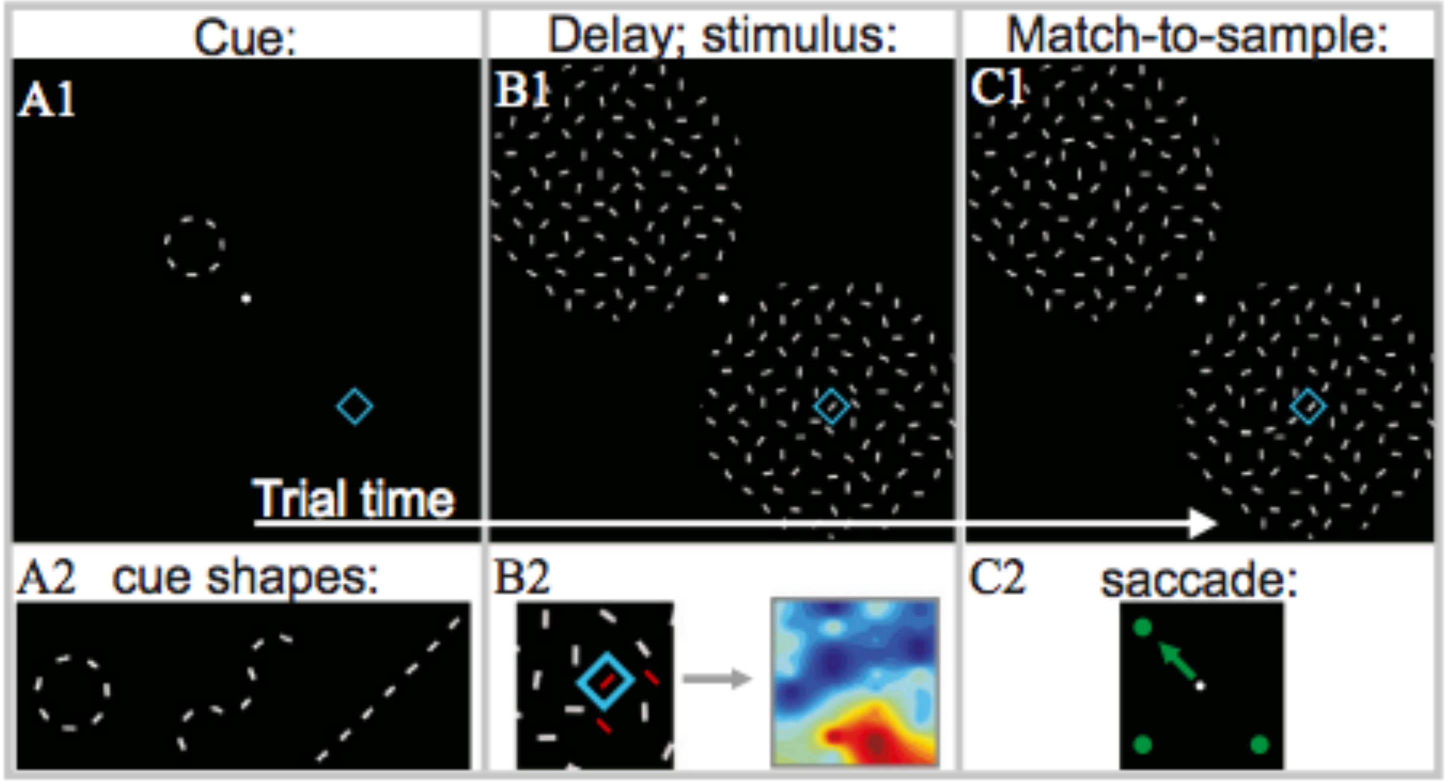
Biological views on function

Adaptive shape processing in primary visual cortex

Justin N. J. McManus[a], Wu Li[b], and Charles D. Gilbert[a,1]

Author Affiliations

Contributed by Charles D. Gilbert, April 18, 2011 (sent for review March 4, 2011)



McManus … Gilbert 2011

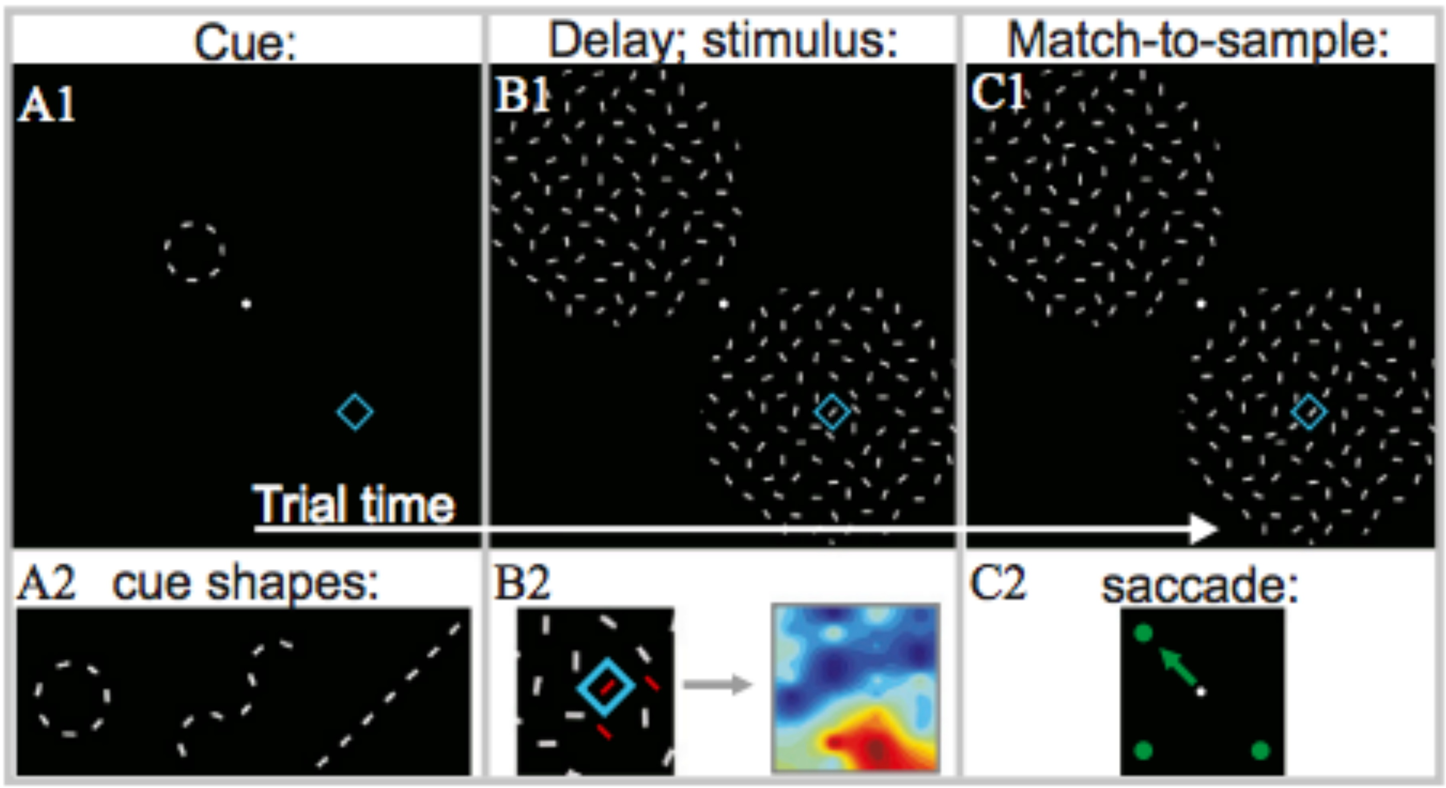
Biological views on function

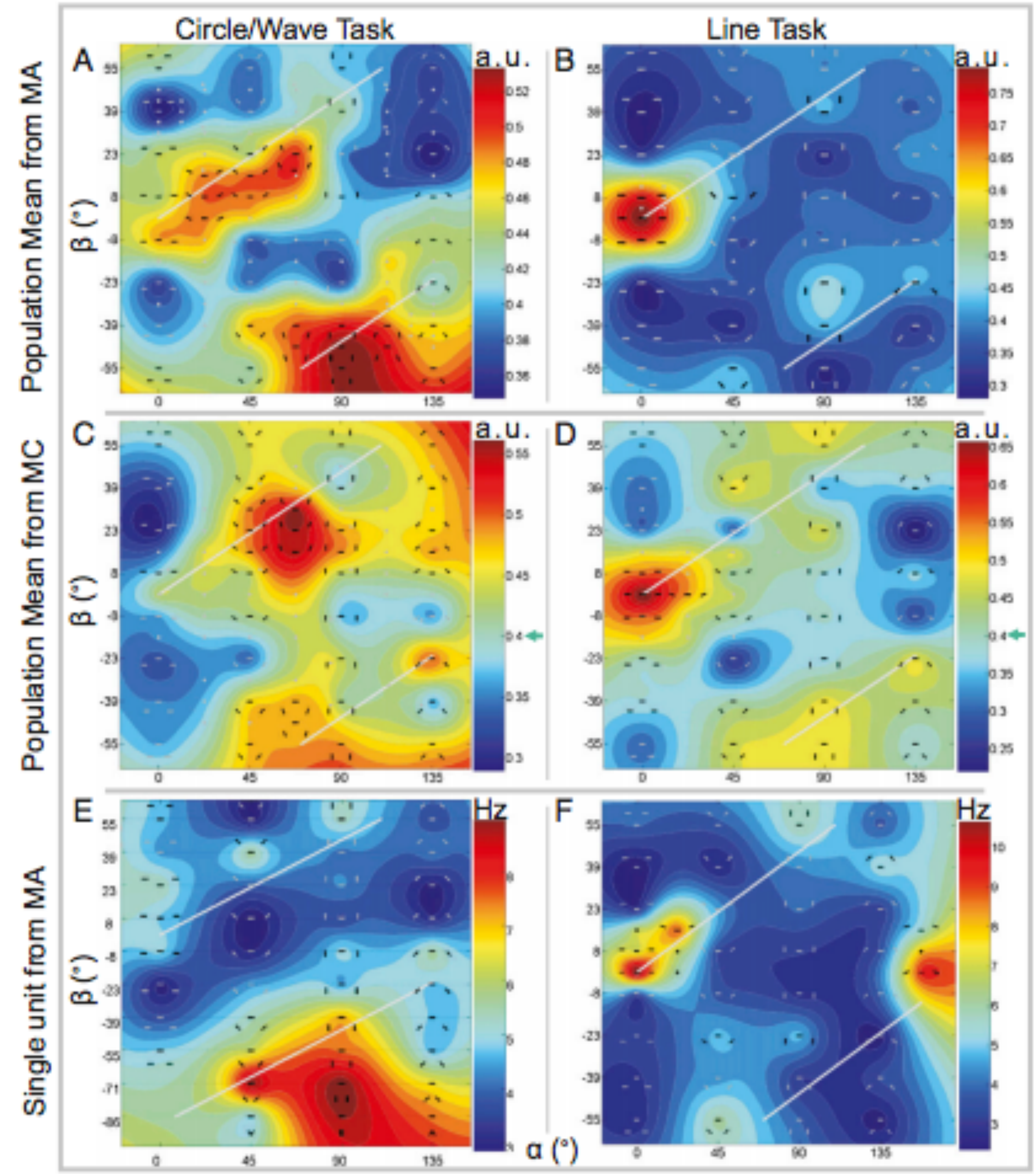
Adaptive shape processing in primary visual cortex

Justin N. J. McManus[a], Wu Li[b], and Charles D. Gilbert[a,1]

Author Affiliations

Contributed by Charles D. Gilbert, April 18, 2011 (sent for review March 4, 2011)





an entirely different mode of selectivity, for circular shapes. The difference between the mean tuning surfaces under the line and circle/wave tasks was statistically significant (monkey A, total number of surfaces, $n = 53$, $P = 4 \times 10^{-5}$; monkey B, $n = 63$, $P = 0.007$; and monkey C, $n = 62$, $P = 0.003$).

McManus … Gilbert 2011

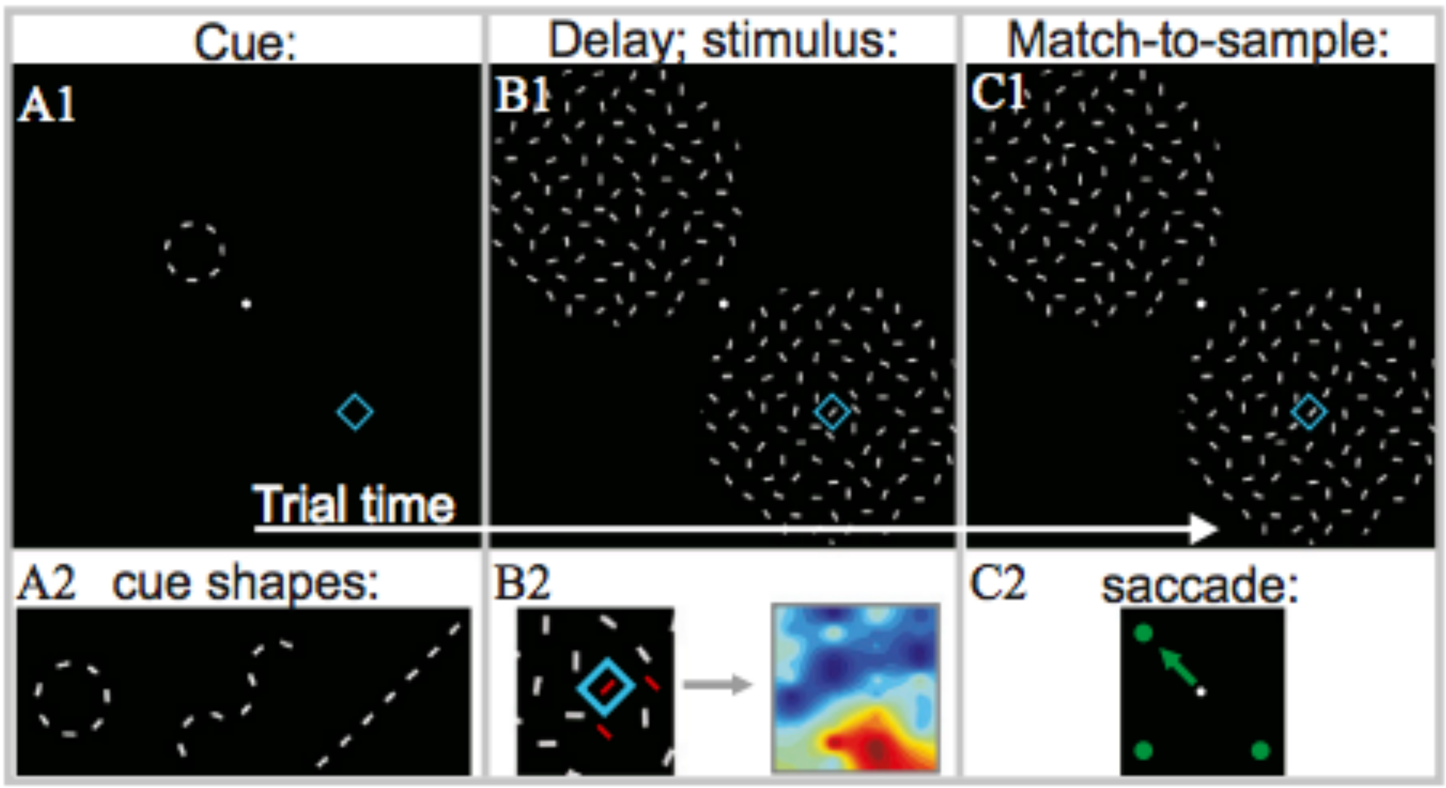
Biological views on function

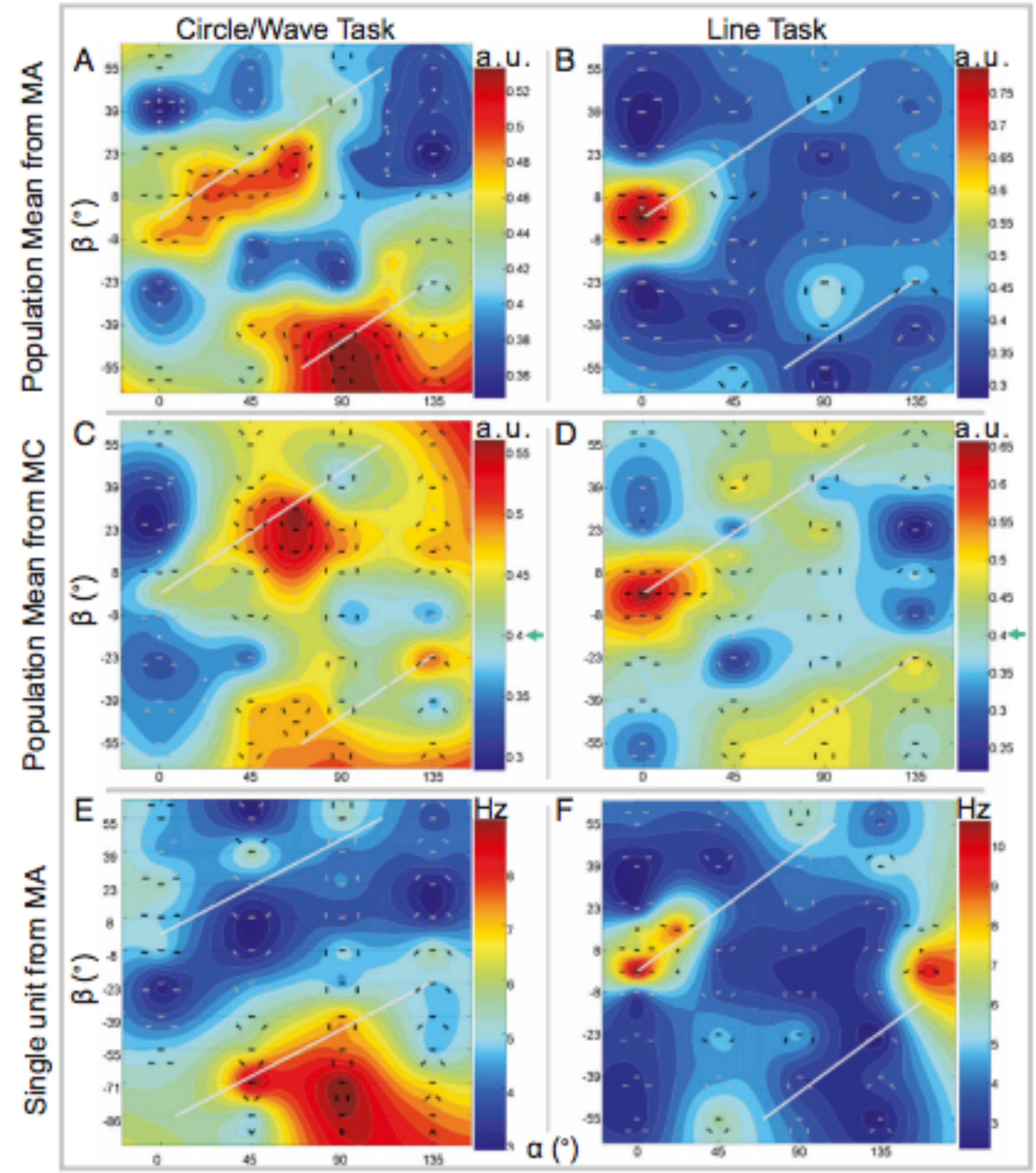
Adaptive shape processing in primary visual cortex

Justin N. J. McManus[a], Wu Li[b], and Charles D. Gilbert[a,1]

Author Affiliations

Contributed by Charles D. Gilbert, April 18, 2011 (sent for review March 4, 2011)

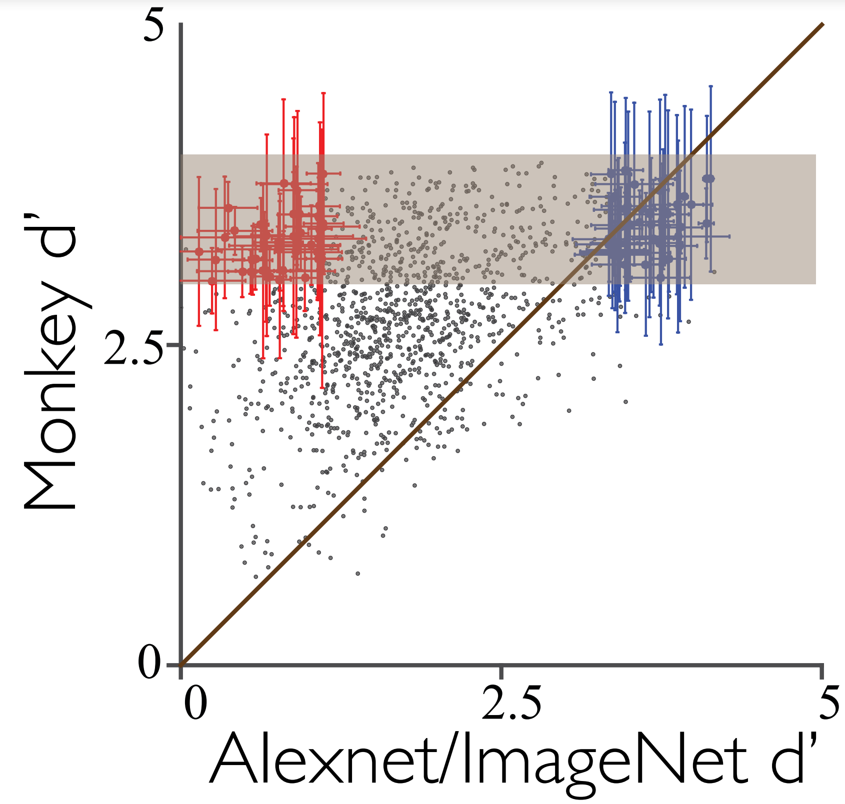




"This process suggests that expectation of an object creates a set of filters that are selective for the object's components and thus, a role of top-down processes in object recognition. The idea is further supported by the transfer of perceptual learning between objects with shared components."

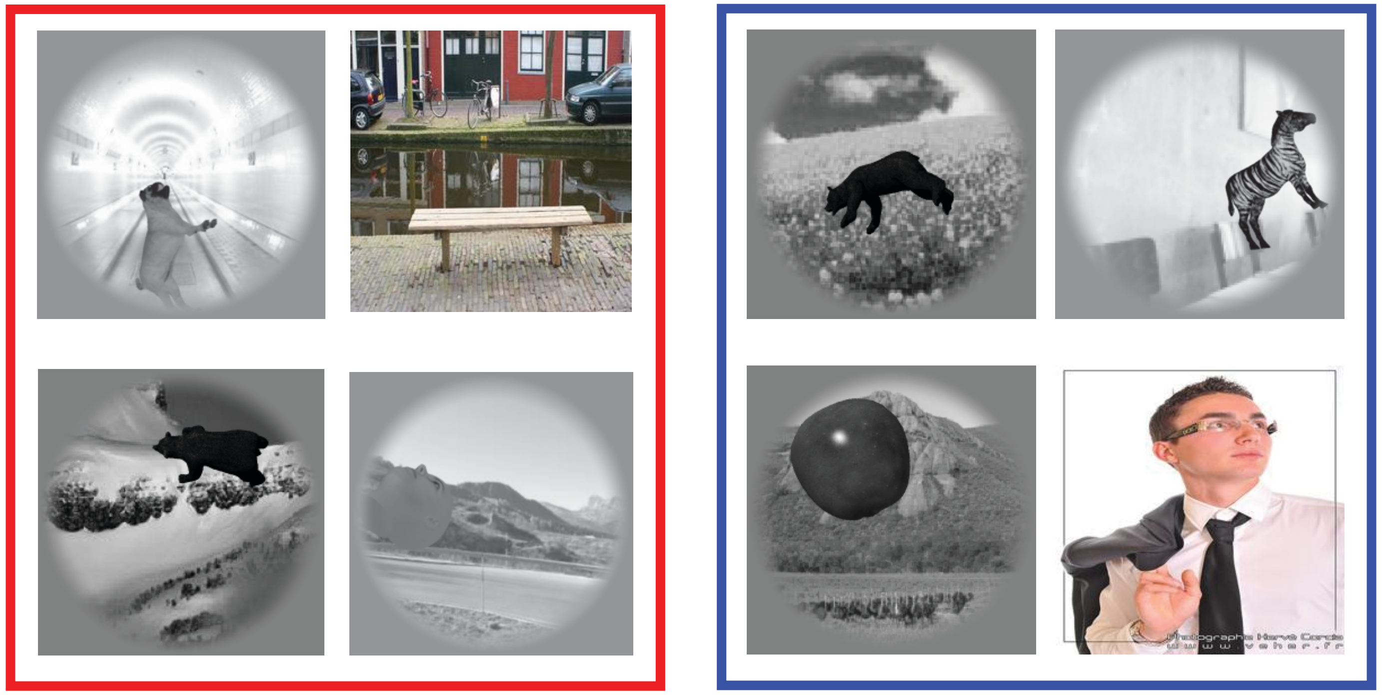
McManus … Gilbert 2011

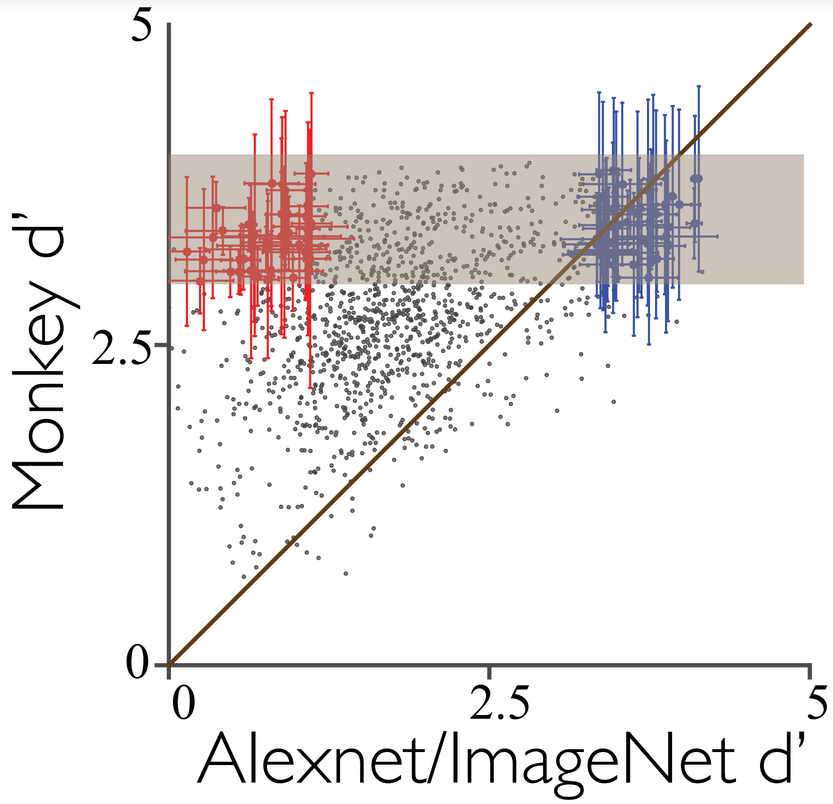
Biological views on function

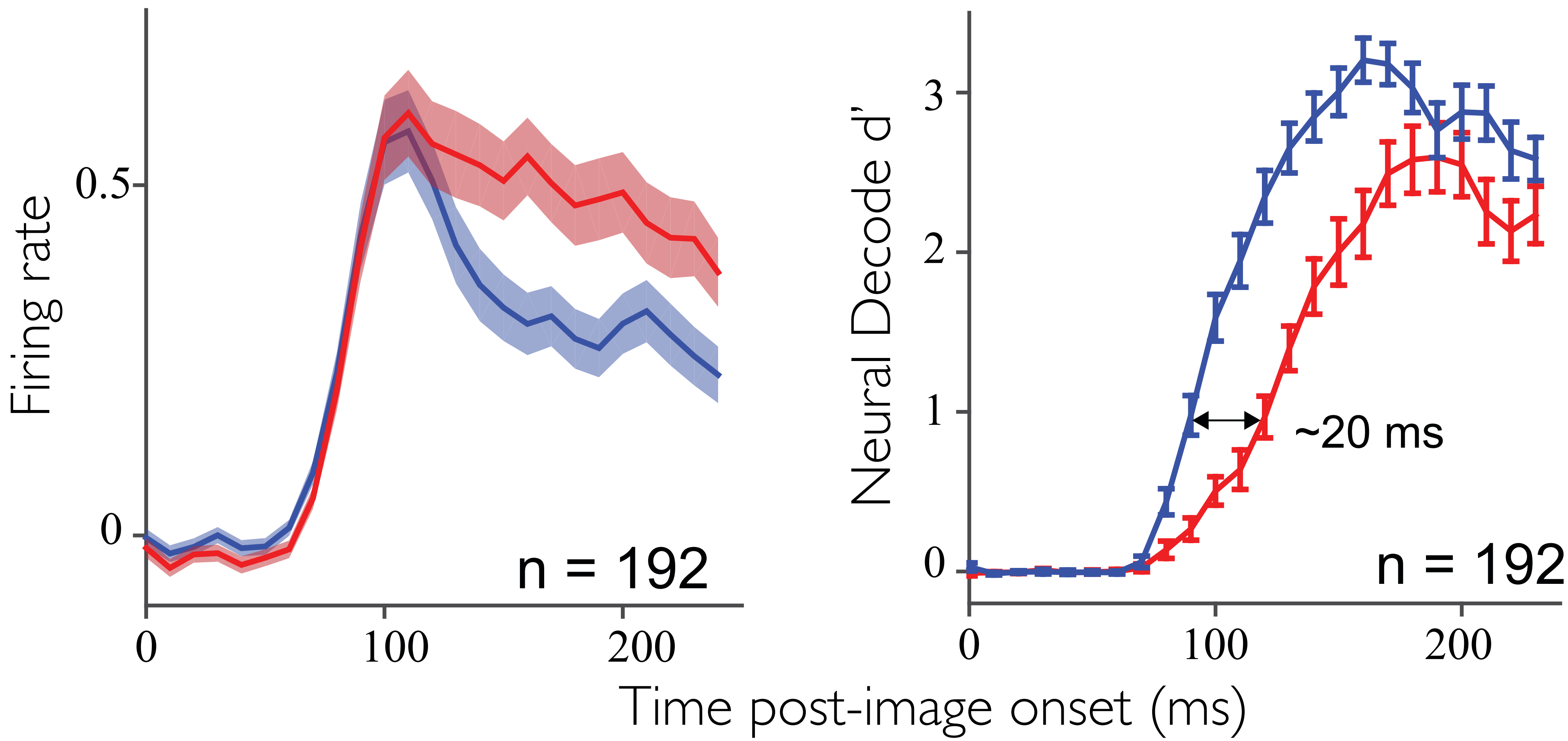


Kar et. al. (2017)

Biological views on function



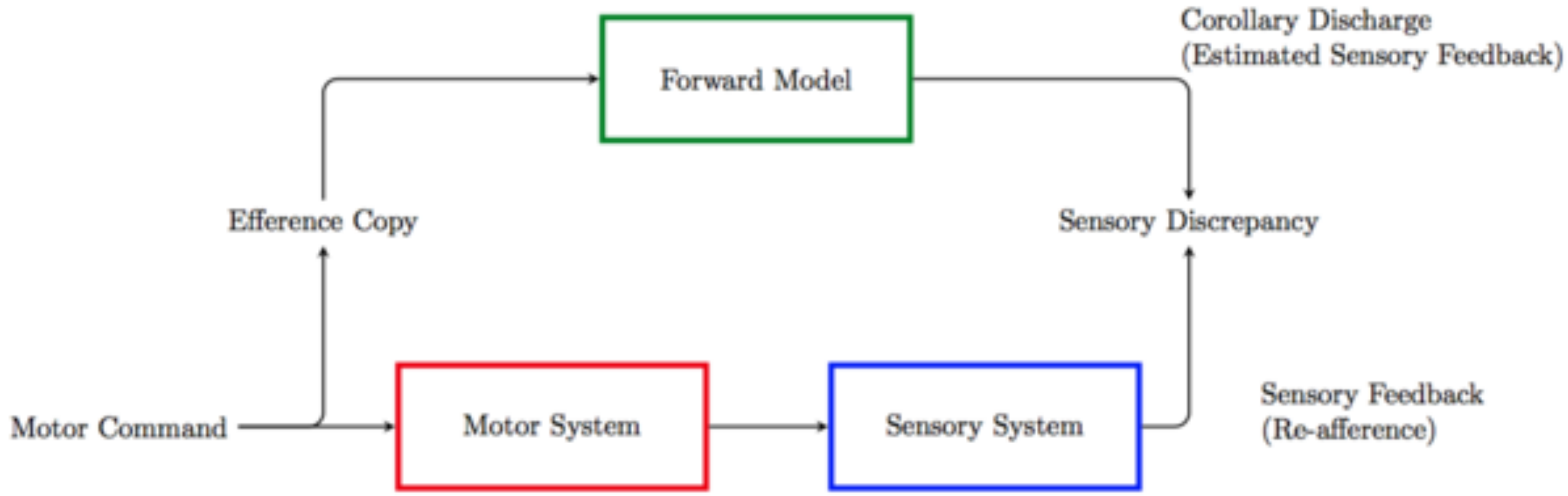




Kar et. al. (2017)

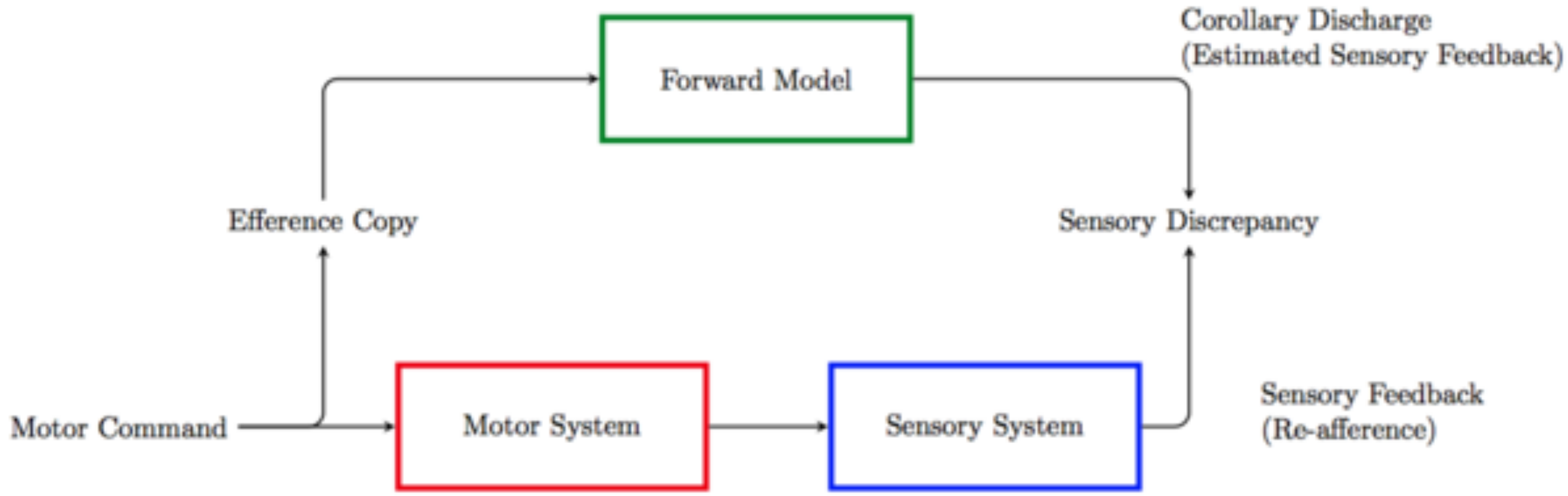
Biological views on function

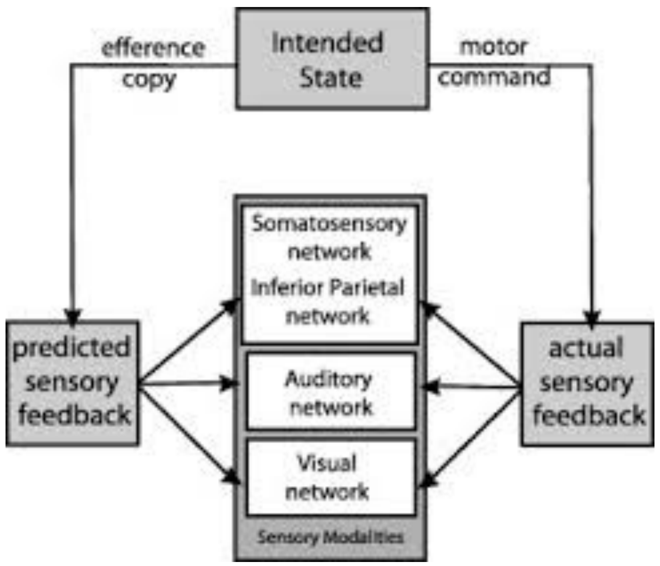
Efference copy = copy of motor instructions, for (e.g.) stability



Biological views on function

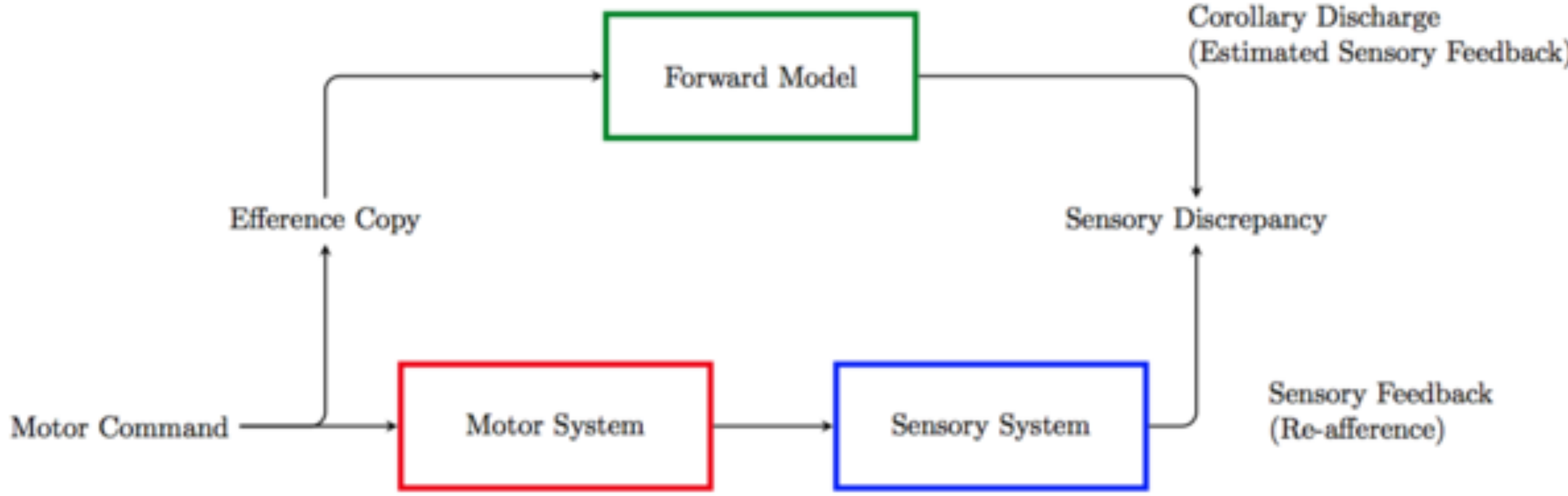
Efference copy = copy of motor instructions, for (e.g.) stability

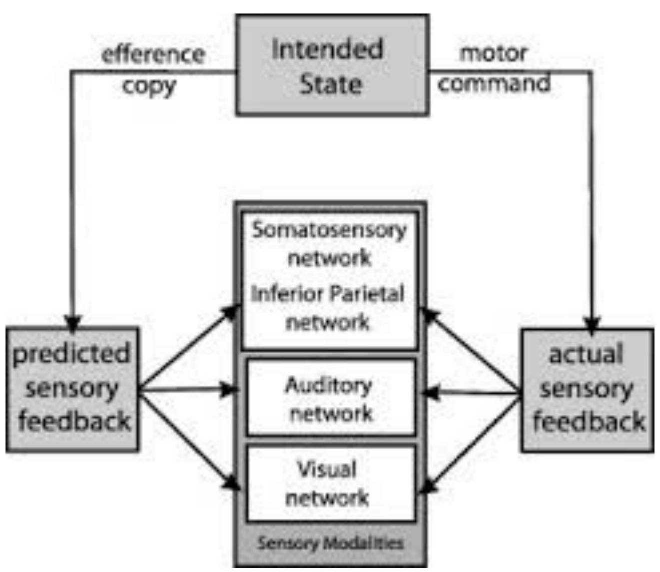


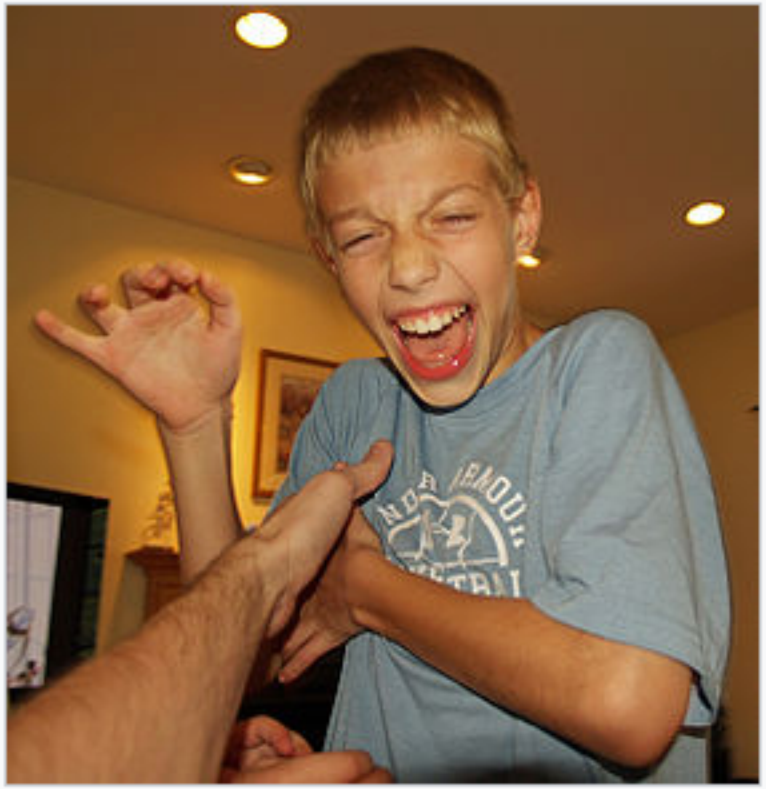


Biological views on function

Efference copy = copy of motor instructions, for (e.g.) stability



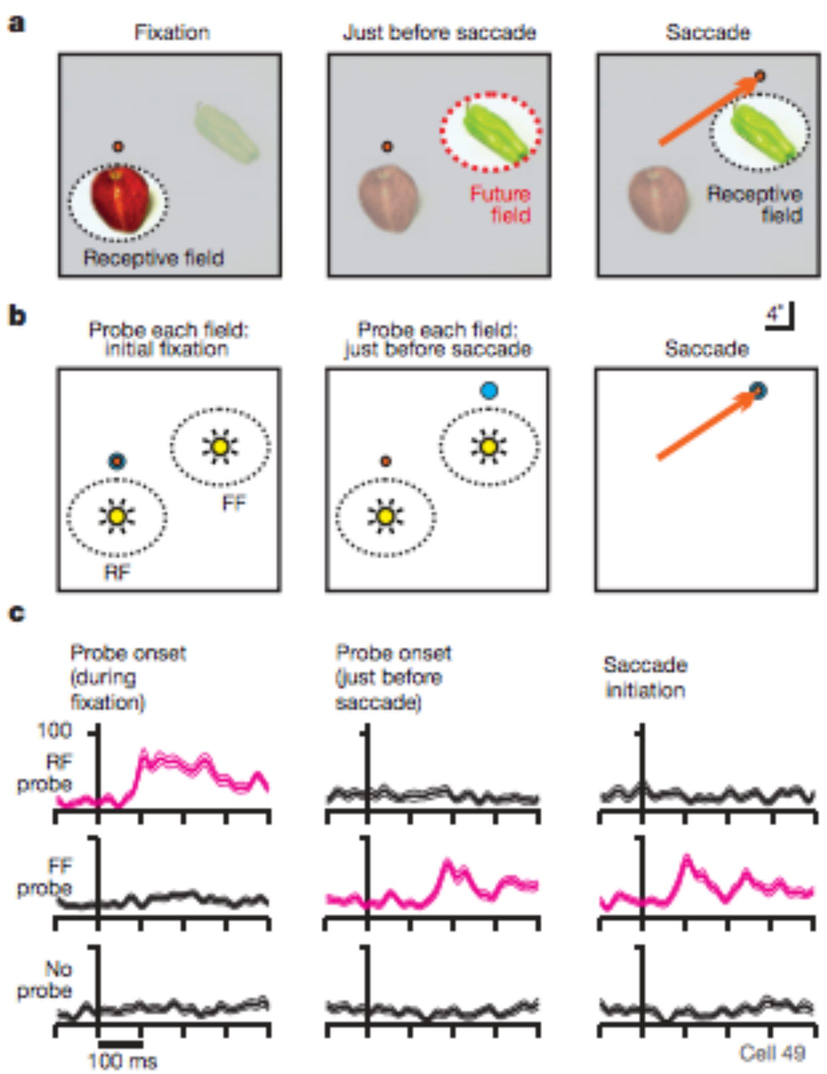




Efference copies are created with our own movement but not those of other people. This is why other people can tickle us (no efference copies of the movements that touch us) but we cannot tickle ourselves (efference copies tell us that we are stimulating ourselves).

Biological views on function

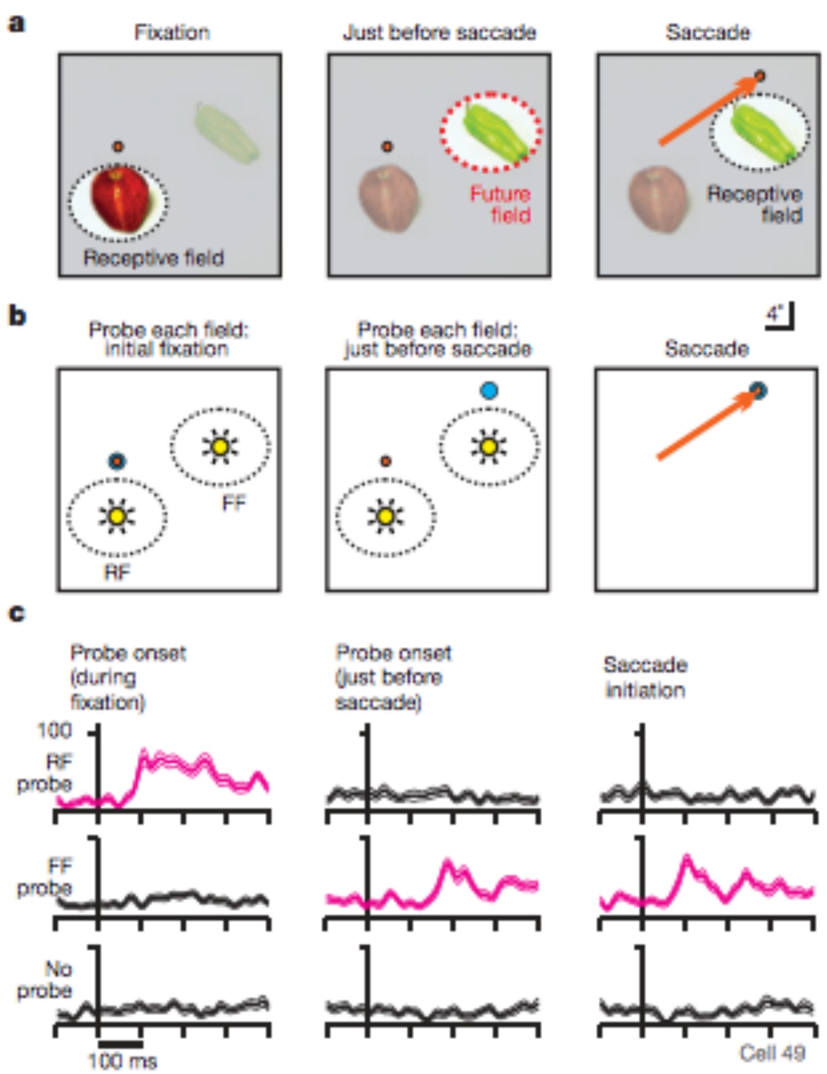
Efference copy = copy of motor instructions, for (e.g.) stability



Summer & Wurtz 2006

Biological views on function

Efference copy = copy of motor instructions, for (e.g.) stability



Superior Colliculus (SC) — "issues motor commands"

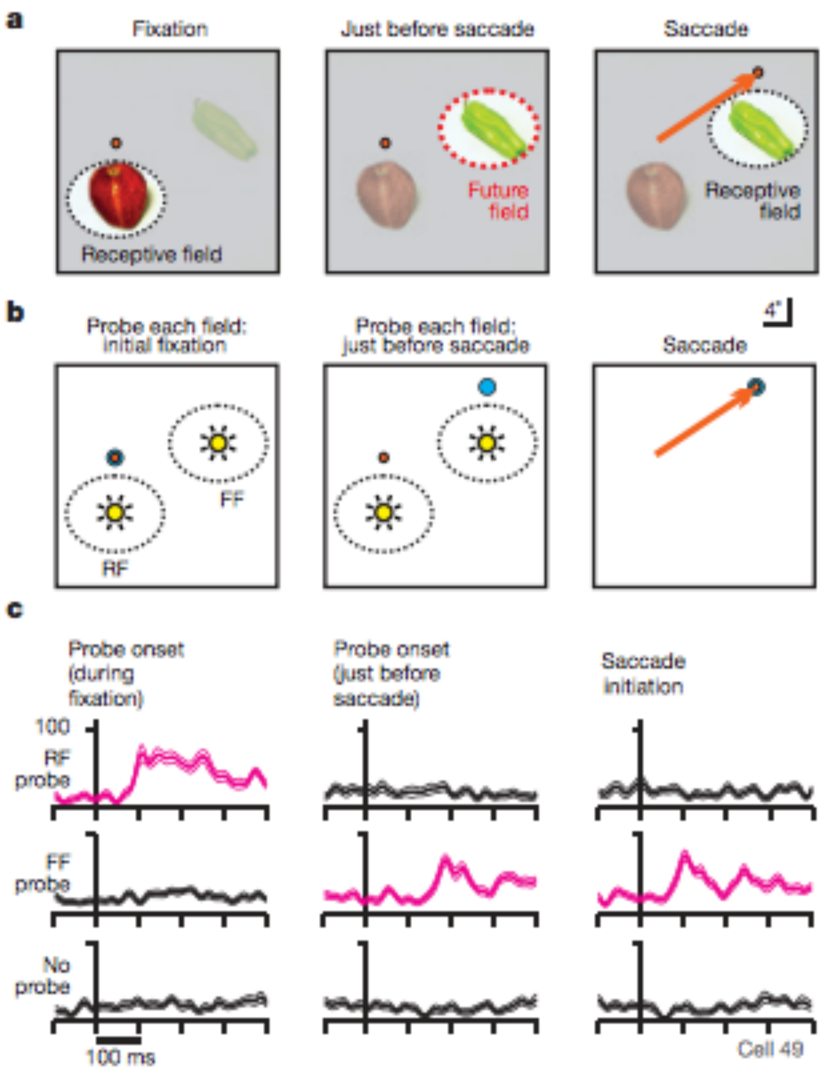
Medial Dorsal (MD) of thalamus — "routing"

Frontal Eye Field (FEF) — moves the eyes

Summer & Wurtz 2006

Biological views on function

Efference copy = copy of motor instructions, for (e.g.) stability



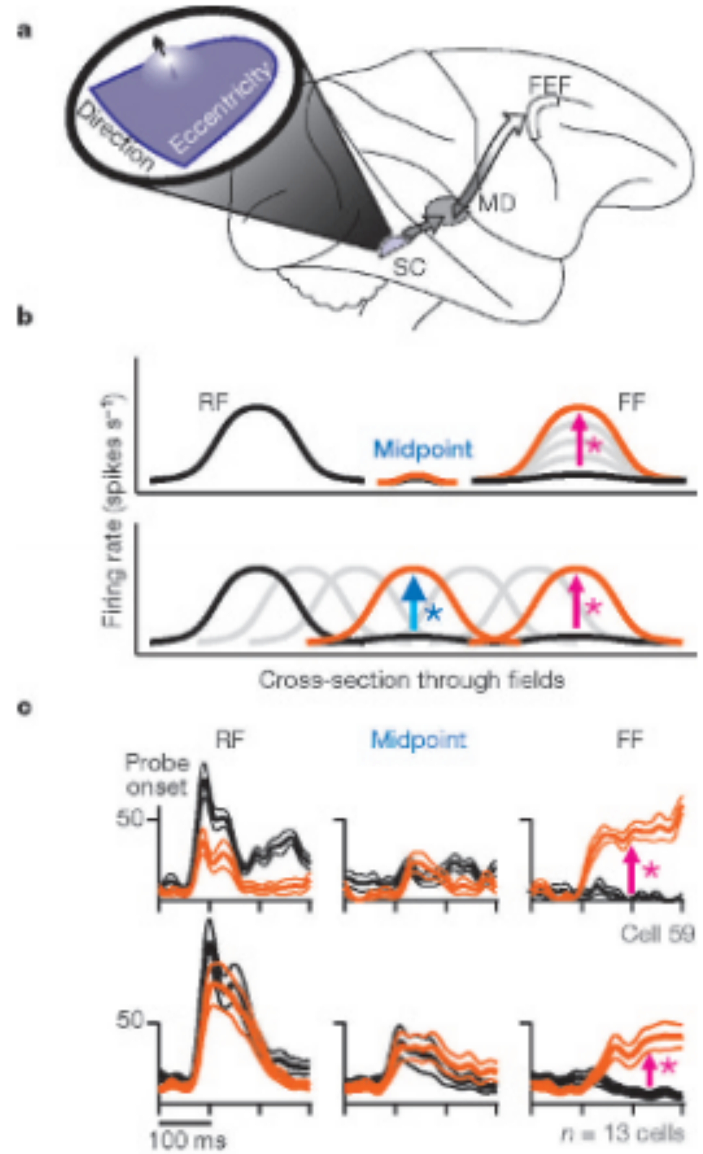


Figure 2 | The spatial properties of shifting RFs are predicted by corollary discharge from the SC–MD–FEF pathway. a, The corollary discharge arises

Superior Colliculus (SC) — "issues motor commands"

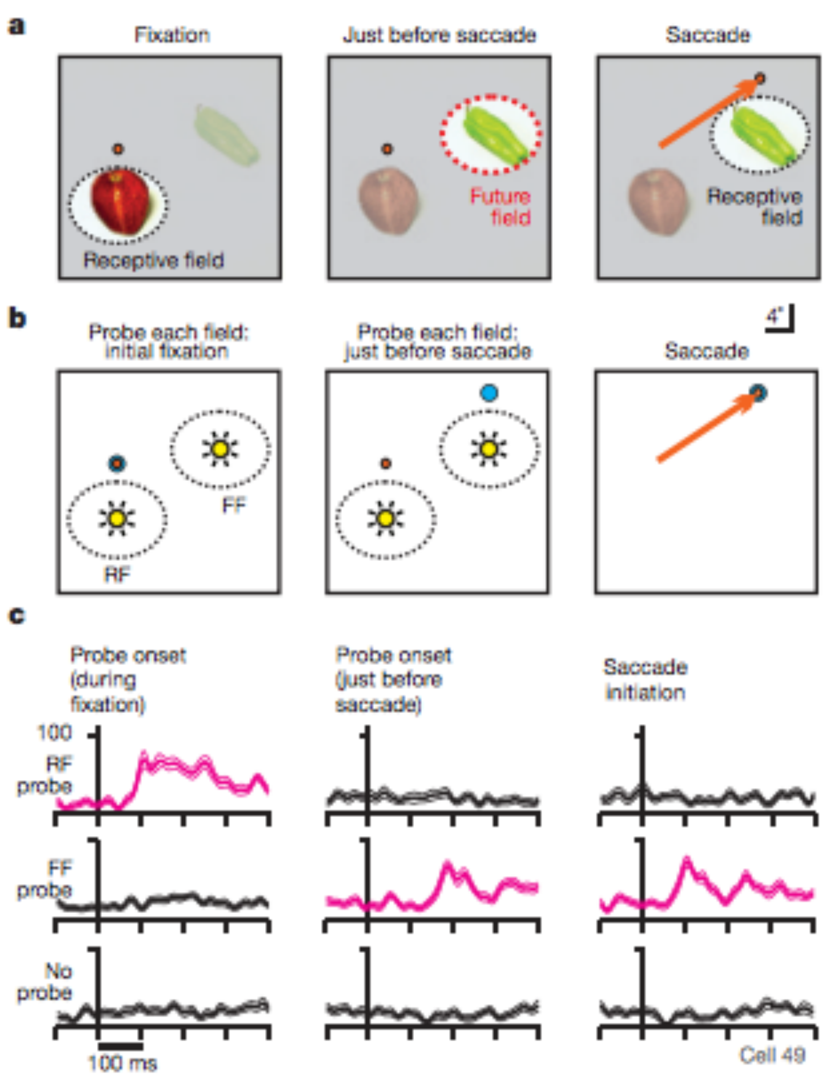
Medial Dorsal (MD) of thalamus — "routing"

Frontal Eye Field (FEF) — moves the eyes

Summer & Wurtz 2006

Biological views on function

Efference copy = copy of motor instructions, for (e.g.) stability



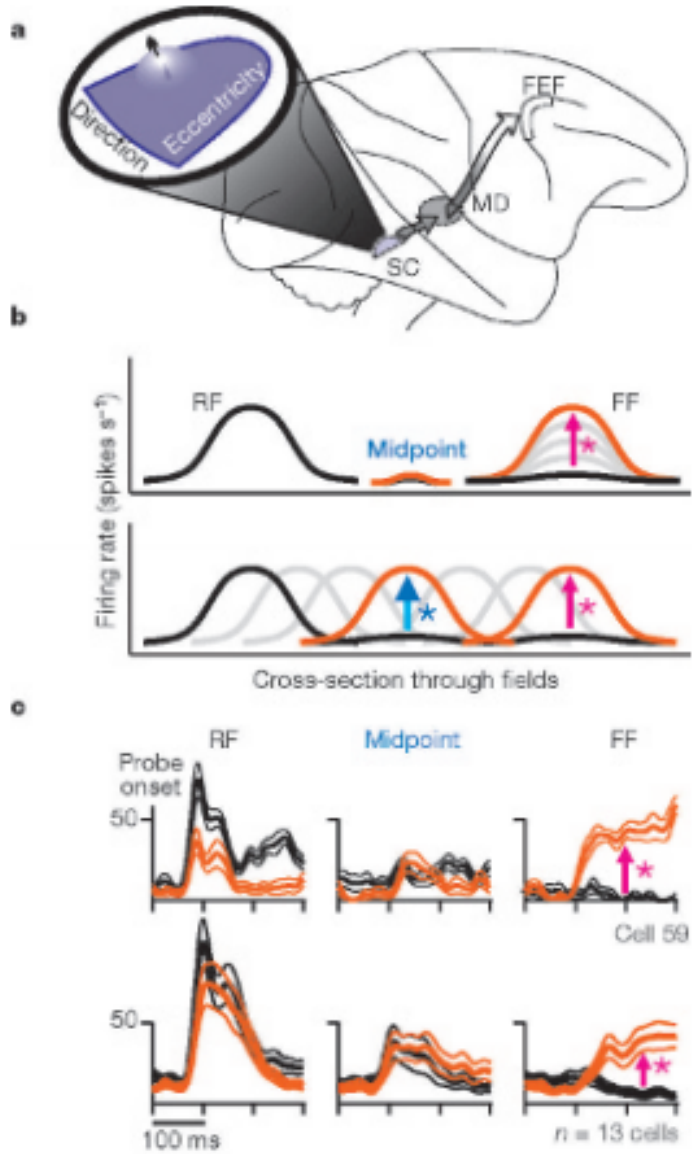


Figure 2 | The spatial properties of shifting RFs are predicted by corollary discharge from the SC–MD–FEF pathway. a, The corollary discharge arises

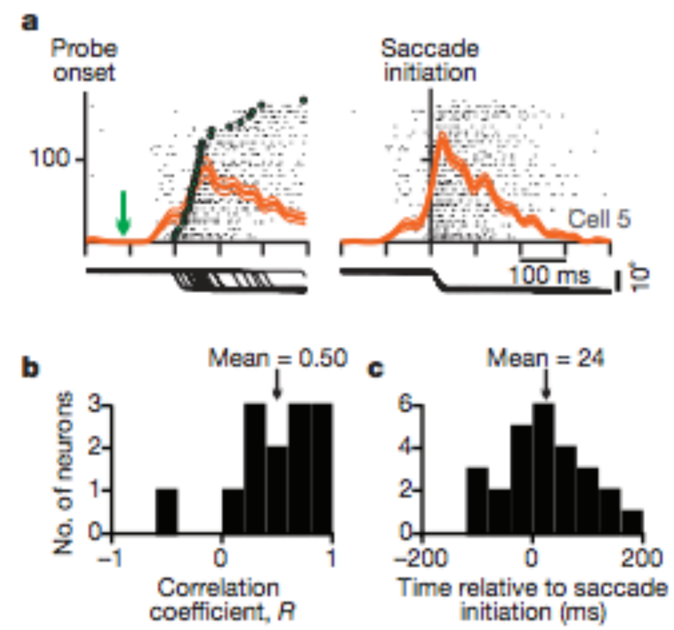


Figure 3 | The temporal properties of shifting RFs are predicted by corollary discharge from the SC–MD–FEF pathway. a, Our hypothesis

Superior Colliculus (SC) — "issues motor commands"

Medial Dorsal (MD) of thalamus — "routing"

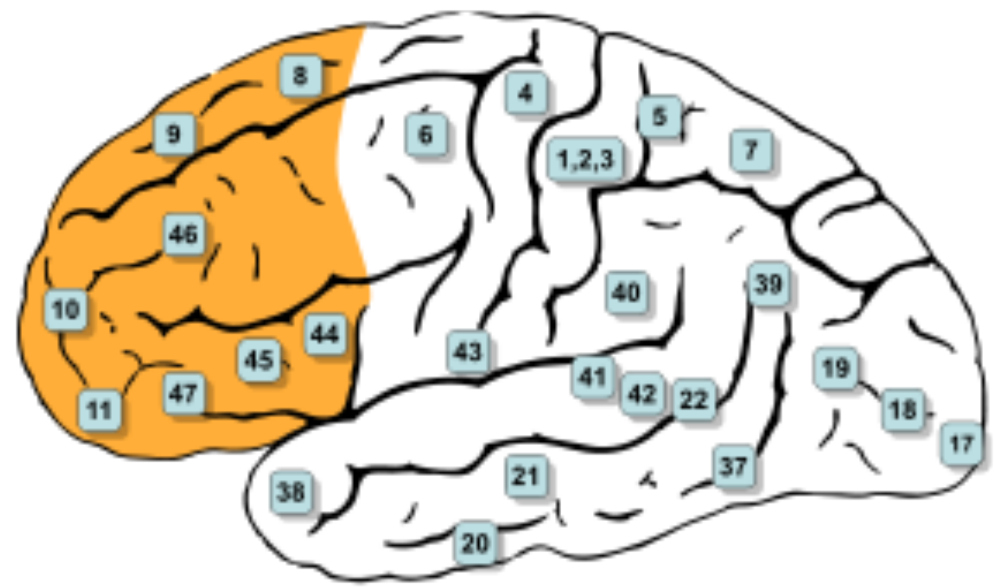
Frontal Eye Field (FEF) — moves the eyes

Summer & Wurtz 2006

Biological views on function

Top-down signal from prefrontal cortex in executive control of memory retrieval

Hyoe Tomita*, Machiko Ohbayashi*, Kiyoshi Nakahara†,
Isao Hasegawa*† & Yasushi Miyashita*†‡



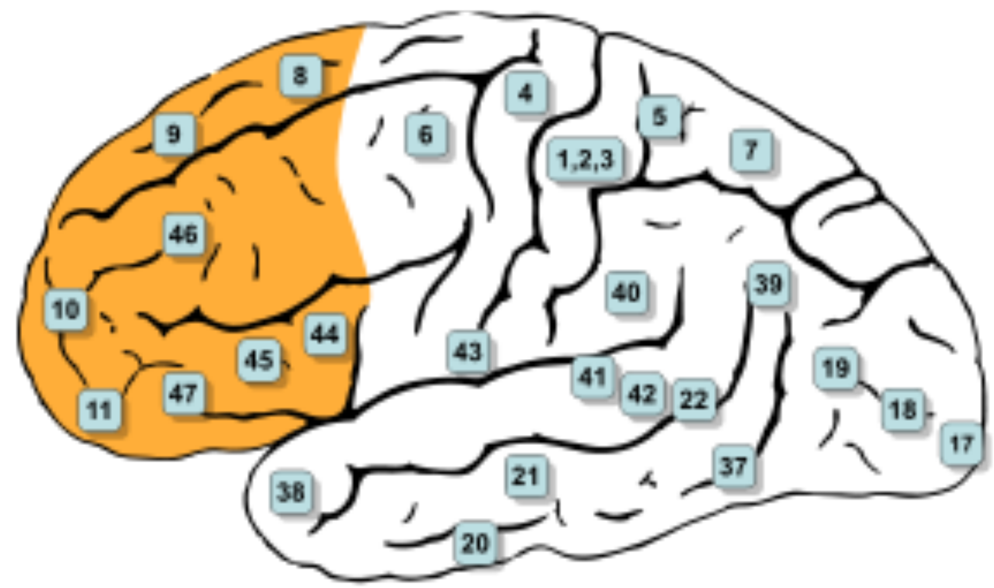
Prefrontal cortex (PFC) ~ “executive control”, long-range planning, decision making, task switching

short-term memory

Biological views on function

Top-down signal from prefrontal cortex in executive control of memory retrieval

Hyoe Tomita*, Machiko Ohbayashi*, Kiyoshi Nakahara†, Isao Hasegawa*† & Yasushi Miyashita*†‡



Prefrontal cortex (PFC) ~ “executive control”, long-range planning, decision making, task switching

short-term memory

Mixture of memory, task (“executive control”), and prediction

“Feedback projections from prefrontal cortex to the posterior association cortex appear to serve the executive control of voluntary recall.”

Biological views on function

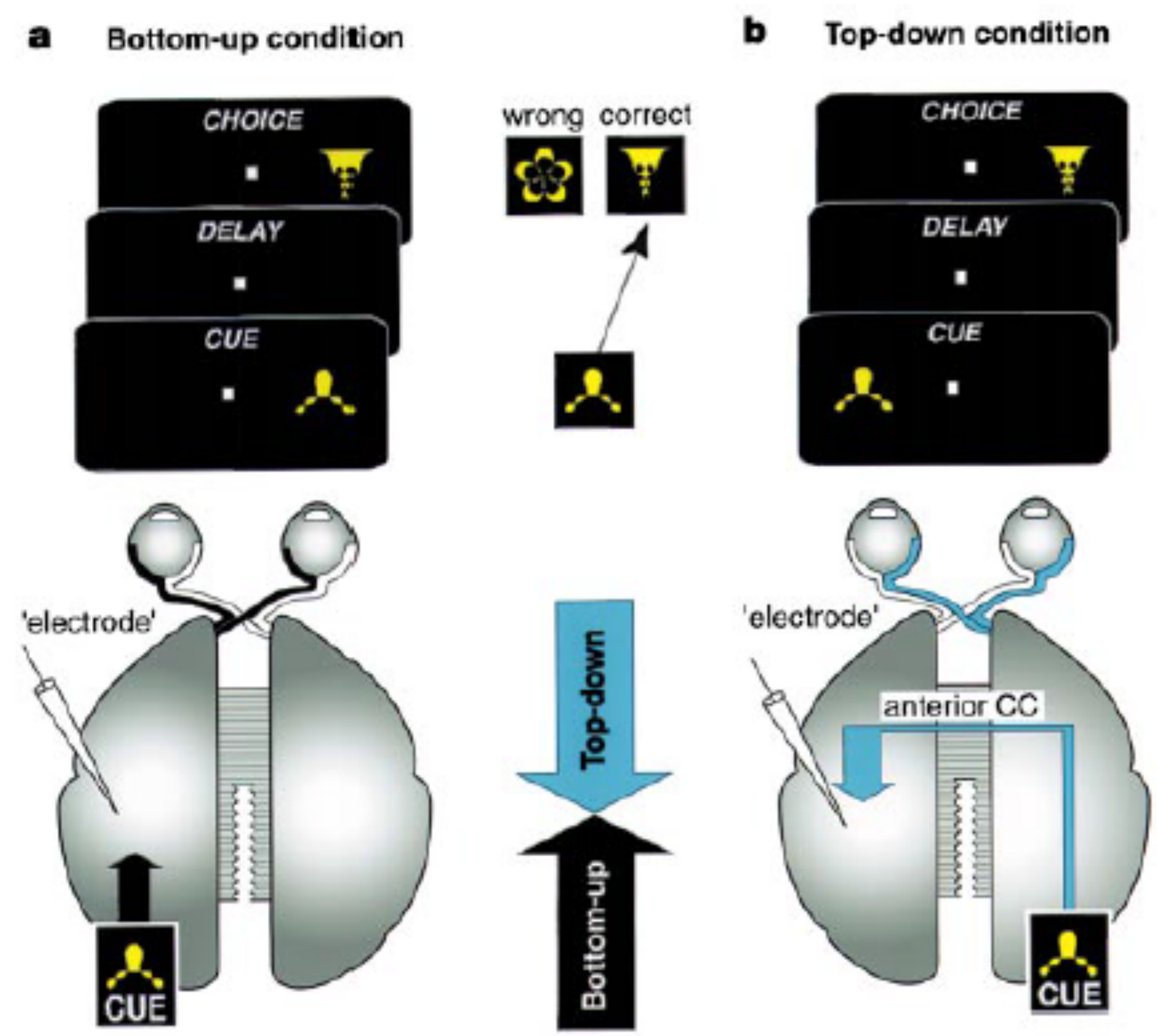
Top-down signal from prefrontal cortex in executive control of memory retrieval

Hyoe Tomita*, Machiko Ohbayashi*, Kiyoshi Nakahara†, Isao Hasegawa*† & Yasushi Miyashita*†‡

Mixture of memory, task (“executive control”), and prediction

“Feedback projections from prefrontal cortex to the posterior association cortex appear to serve the executive control of voluntary recall.”

“Split-brain paradigm” — transection of posterior corpus callosum — IT neurons in one hemisphere are activated by direct bottom-up inputs only in the contralateral hemifield, but not when the inputs are in the ipsilateral hemifield.



Biological views on function

Top-down signal from prefrontal cortex in executive control of memory retrieval

Hyoe Tomita*, Machiko Ohbayashi*, Kiyoshi Nakahara†, Isao Hasegawa*† & Yasushi Miyashita*†‡

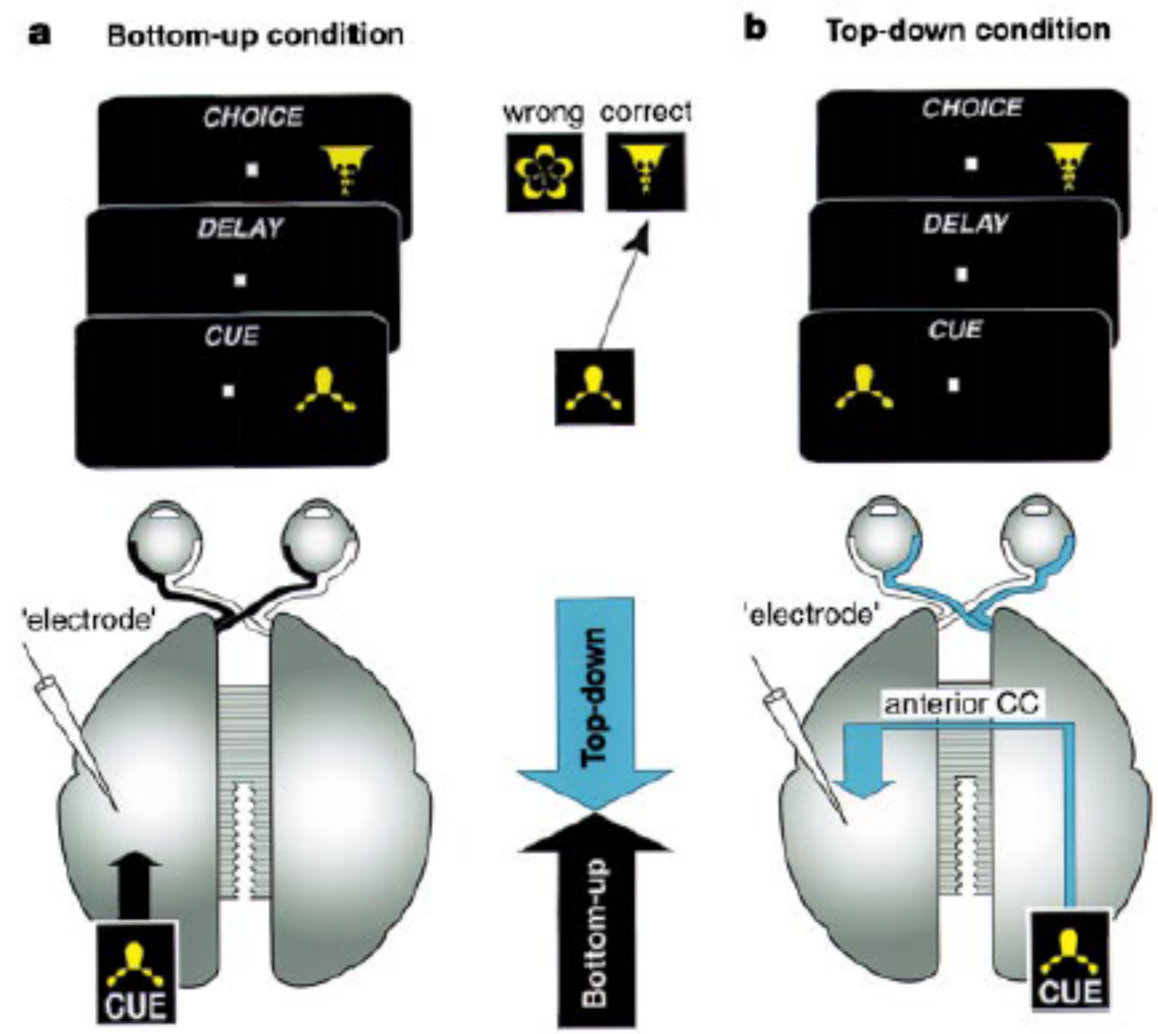
Mixture of memory, task (“executive control”), and prediction

“Feedback projections from prefrontal cortex to the posterior association cortex appear to serve the executive control of voluntary recall.”

“Split-brain paradigm” — transection of posterior corpus callosum — IT neurons in one hemisphere are activated by direct bottom-up inputs only in the contralateral hemifield, but not when the inputs are in the ipsilateral hemifield.

Ipsilateral presentation *still* activated IT neurons, but later than contralateral.

Neuron’s pattern of responses across stimuli similar regardless of ipsi/contra presentation (r = ~0.8)



Biological views on function

Top-down signal from prefrontal cortex in executive control of memory retrieval

Hyoe Tomita*, Machiko Ohbayashi*, Kiyoshi Nakahara†, Isao Hasegawa*† & Yasushi Miyashita*†‡

Mixture of memory, task (“executive control”), and prediction

“Feedback projections from prefrontal cortex to the posterior association cortex appear to serve the executive control of voluntary recall.”

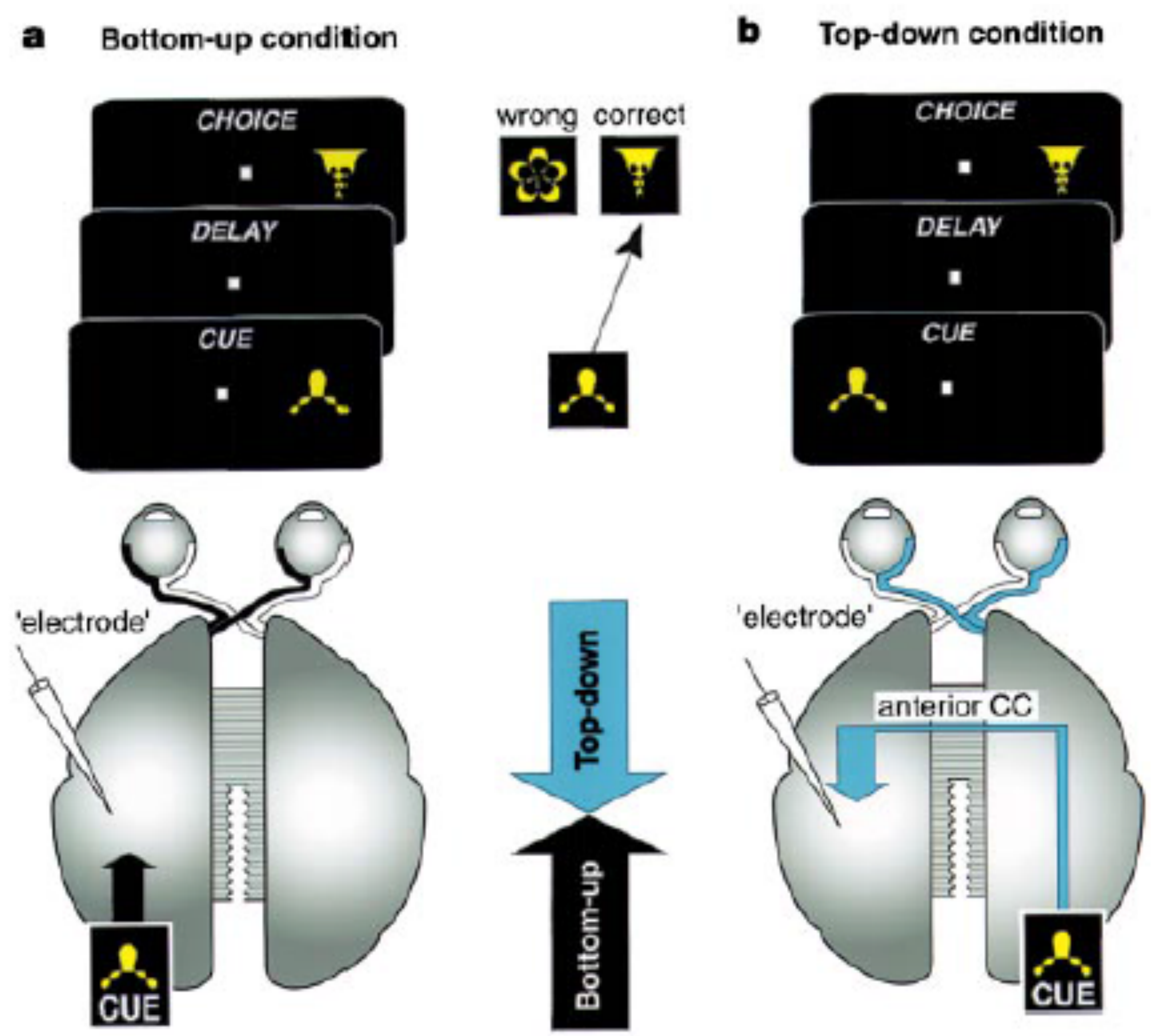
“Split-brain paradigm” — transection of posterior corpus callosum — IT neurons in one hemisphere are activated by direct bottom-up inputs only in the contralateral hemifield, but not when the inputs are in the ipsilateral hemifield.

Ipsilateral presentation *still* activated IT neurons, but later than contralateral.

Neuron’s pattern of responses across stimuli similar regardless of ipsi/contra presentation (r = ~0.8)

No such transfer in *full* split.

Pair associated test indicates prospective information from PFC sent to IT.



Biological views on function

Attention

Neuron

Volume 61, Issue 2, 29 January 2009, Pages 168-185

Review

The Normalization Model of Attention

John H. Reynolds [1], David J. Heeger [2]

Show more

https://doi.org/10.1016/j.neuron.2009.01.002

Get rights and content

Biological views on function

Neuron

CellPress

Volume 61, Issue 2, 29 January 2009, Pages 168-185

Review

The Normalization Model of Attention

John H. Reynolds [1], David J. Heeger [2]

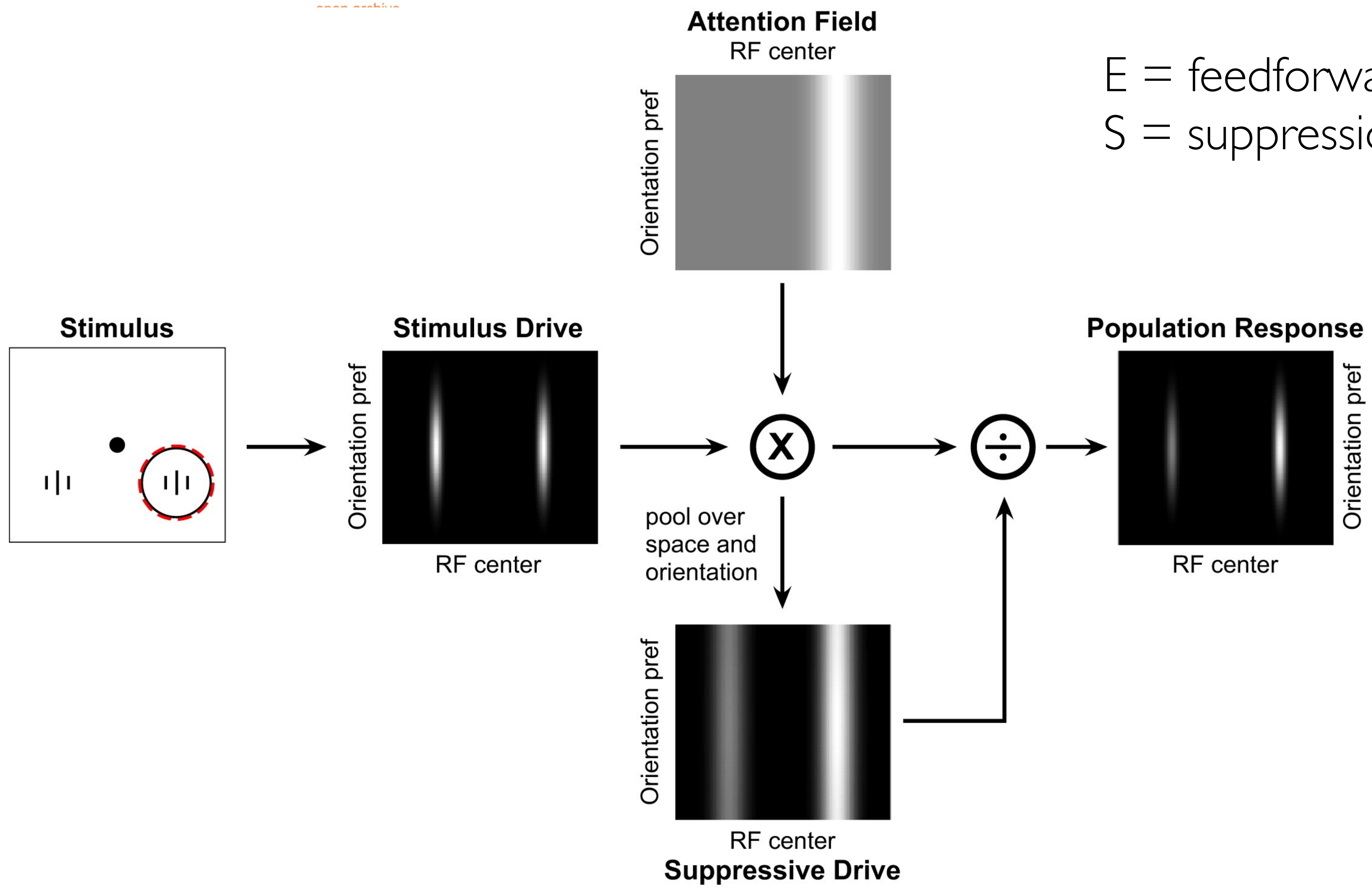
⊞ Show more

https://doi.org/10.1016/j.neuron.2009.01.002

$$R(x,\theta) = \mathbf{ReLu}_T\left[\frac{E(x,\theta)}{S(x,\theta)+\sigma}\right]$$

E = feedforward input

S = suppression



Biological views on function

Neuron

CellPress

Volume 61, Issue 2, 29 January 2009, Pages 168-185

Review

The Normalization Model of Attention

John H. Reynolds [1], David J. Heeger [2]

⊞ Show more

https://doi.org/10.1016/j.neuron.2009.01.002

Get rights and content

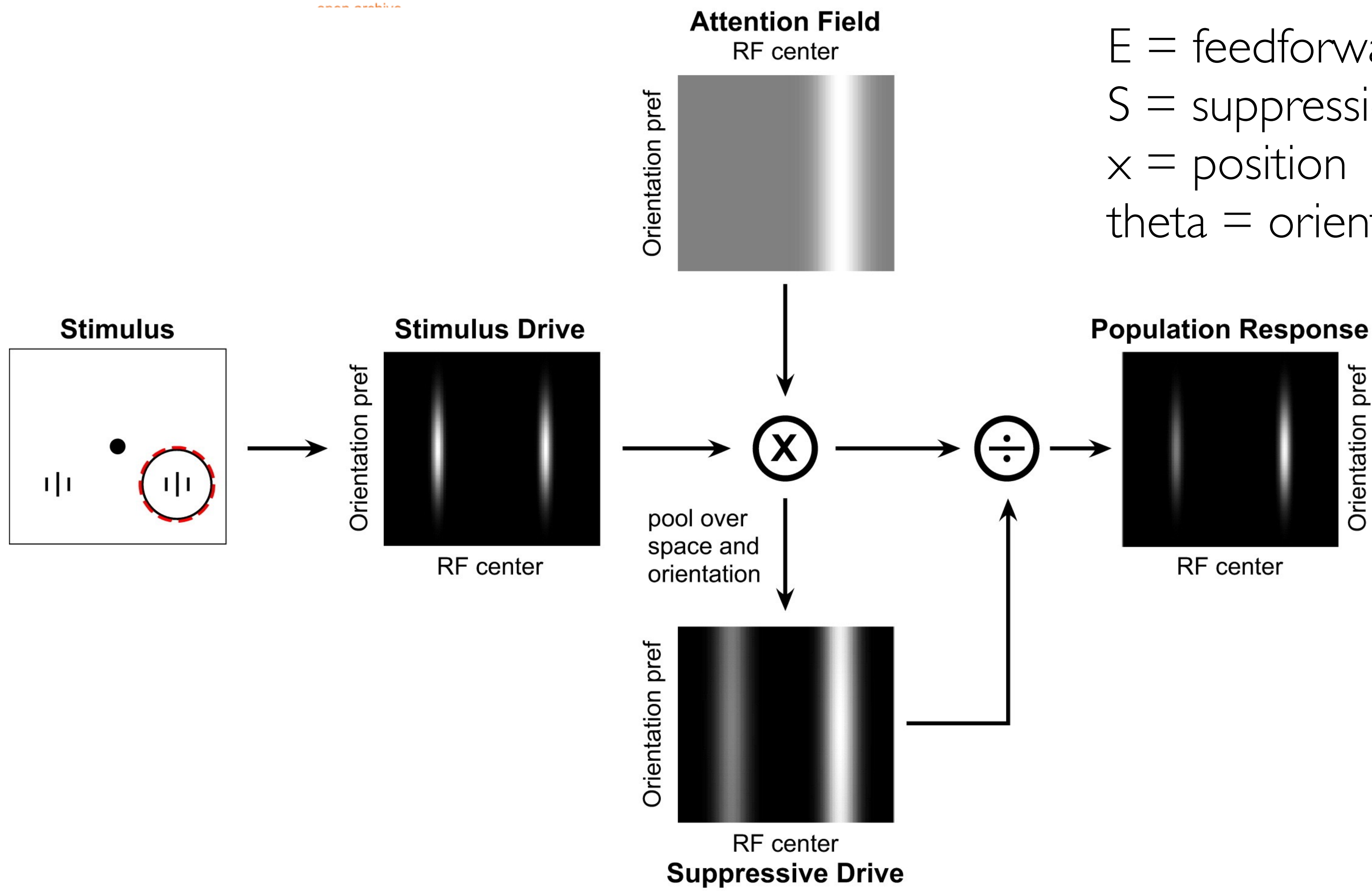
$$R(x,\theta) = \mathbf{ReLu}_T\left[\frac{E(x,\theta)}{\mathbf{Conv}_{s(x,\theta)}[E(x,\theta)] + \sigma}\right]$$

E = feedforward input

S = suppression

x = position

theta = orientation



Biological views on function

Neuron

CellPress

Volume 61, Issue 2, 29 January 2009, Pages 168-185

Review

The Normalization Model of Attention

John H. Reynolds [1], David J. Heeger [2]

Show more

https://doi.org/10.1016/j.neuron.2009.01.002

Get rights and content

$$R(x,\theta) = \mathbf{ReLu}_T\left[\frac{A(x,\theta)E(x,\theta)}{\mathbf{Conv}_{s(x,\theta)}[A(x,\theta)E(x,\theta)] + \sigma}\right]$$

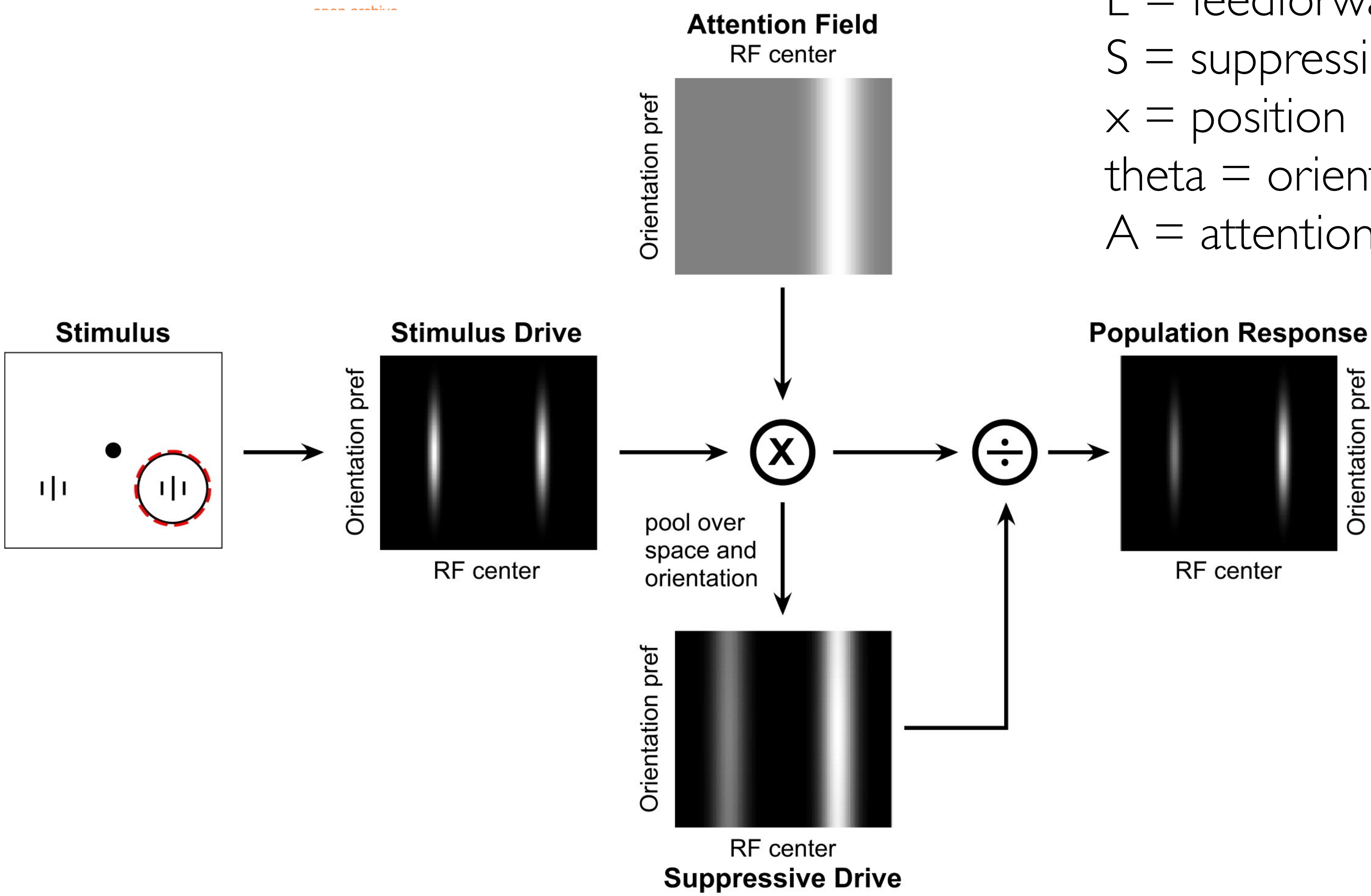
E = feedforward input

S = suppression

x = position

theta = orientation

A = attention field



—> implemented as *equilibrium* of simple recurrent circuit (Heeger 1993)

Biological views on function

Top-down influence in early visual processing
A Bayesian perspective

Tai Sing Lee

Center for the Neural Basis of Cognition
Department of Computer Science
Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.
Department of Neuroscience
University of Pittsburgh, Pittsburgh, PA 15213, U.S.A.

$$P(S_i \mid E, H) = \frac{P(E \mid S_i, H) P(S_i \mid H)}{P(E \mid H)}$$

S_i = scene i

E = evidence

H = prior information

Biological views on function

Top-down influence in early visual processing
A Bayesian perspective

Tai Sing Lee

Center for the Neural Basis of Cognition
Department of Computer Science
Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.
Department of Neuroscience
University of Pittsburgh, Pittsburgh, PA 15213, U.S.A.

Bayesian interaction of two brain areas

$$P(S_i \mid E, H) = \frac{P(E \mid S_i, H)P(S_i \mid H)}{P(E \mid H)}$$

S_i = scene i output of V1

E = evidence finished to V1 by retina

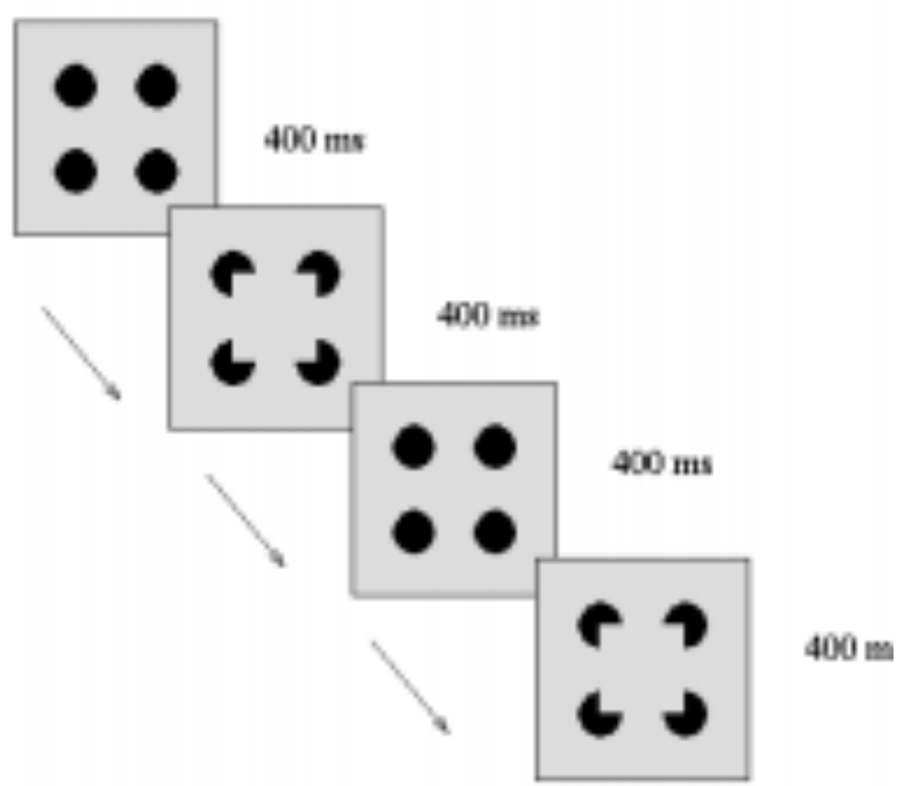
H = prior information generated by V2

Biological views on function

Top-down influence in early visual processing
A Bayesian perspective

Tai Sing Lee

Center for the Neural Basis of Cognition
Department of Computer Science
Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.
Department of Neuroscience
University of Pittsburgh, Pittsburgh, PA 15213, U.S.A.



<u>Bayesian interaction of two brain areas</u>

$$P(S_i \mid E, H) = \frac{P(E \mid S_i, H)P(S_i \mid H)}{P(E \mid H)}$$

S_i = scene i output of V1

E = evidence finished to V1 by retina

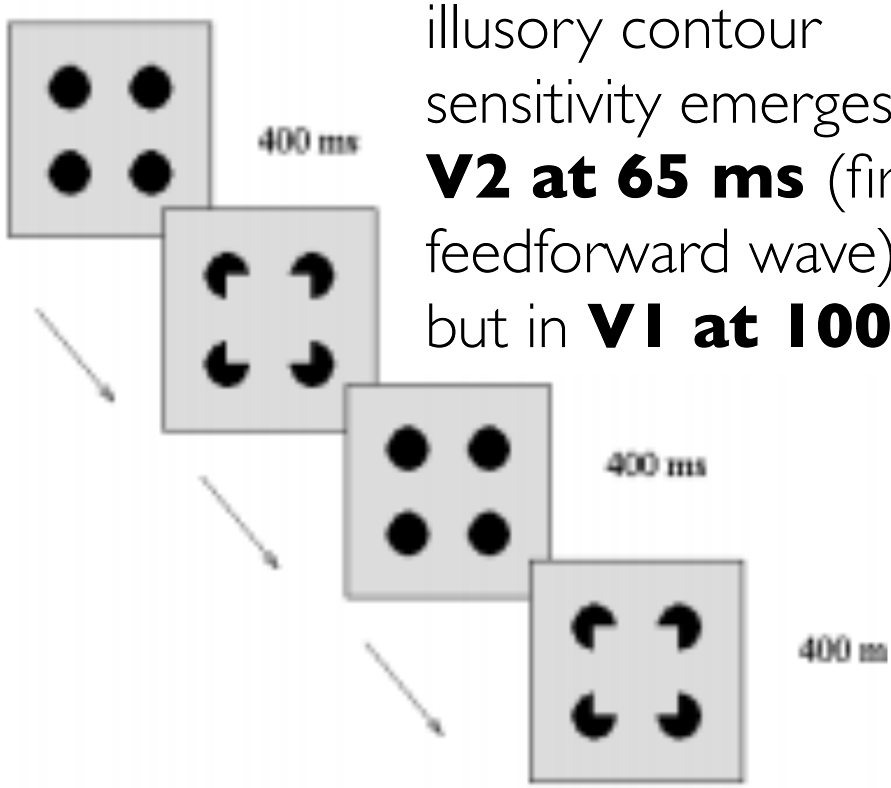
H = prior information generated by V2

Biological views on function

Top-down influence in early visual processing
A Bayesian perspective

Tai Sing Lee

Center for the Neural Basis of Cognition
Department of Computer Science
Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.
Department of Neuroscience
University of Pittsburgh, Pittsburgh, PA 15213, U.S.A.



illusory contour sensitivity emerges in **V2 at 65 ms** (first feedforward wave) but in **V1 at 100ms**

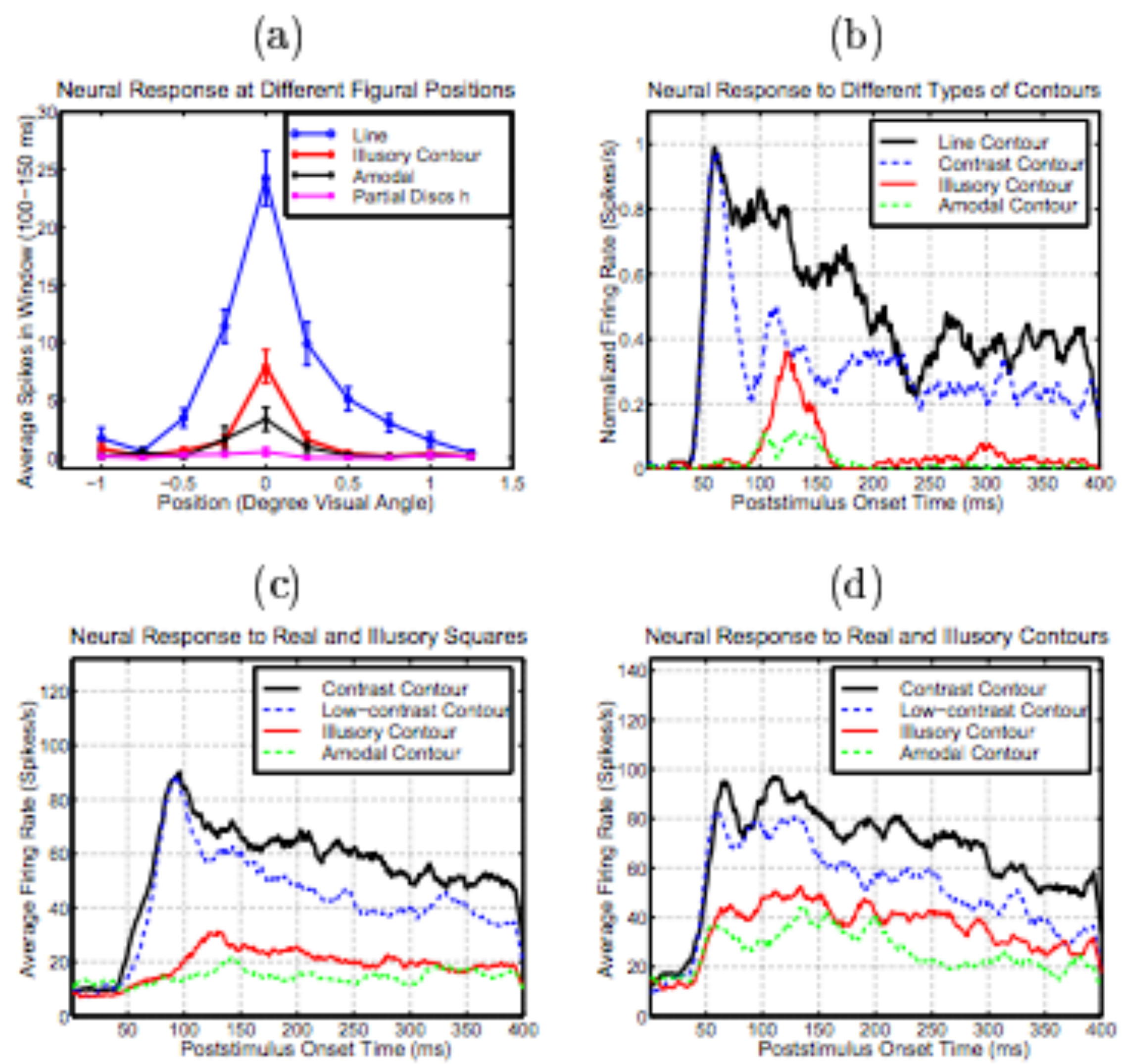
Bayesian interaction of two brain areas

$$P(S_i \mid E, H) = \frac{P(E \mid S_i, H)P(S_i \mid H)}{P(E \mid H)}$$

S_i = scene i output of V1

E = evidence finished to V1 by retina

H = prior information generated by V2



Biological views on function

Task Dependence Executive control

Adaptive Shape processing

Efferent Copy

Memory

Generalized Attention Bayesian inference

Implementing Learning

Models

What and where: A Bayesian inference theory of attention

Sharat Chikkerur *, Thomas Serre, Cheston Tan, Tomaso Poggio

McGovern Institute for Brain Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, United States

$$P(O, L, X^1, \ldots, X^N, I) = P(O)P(L)P(I \mid X^1, \ldots, X^N) \prod_{i=1}^{N} P(X^i \mid L, F^i)P(F^i \mid O)$$

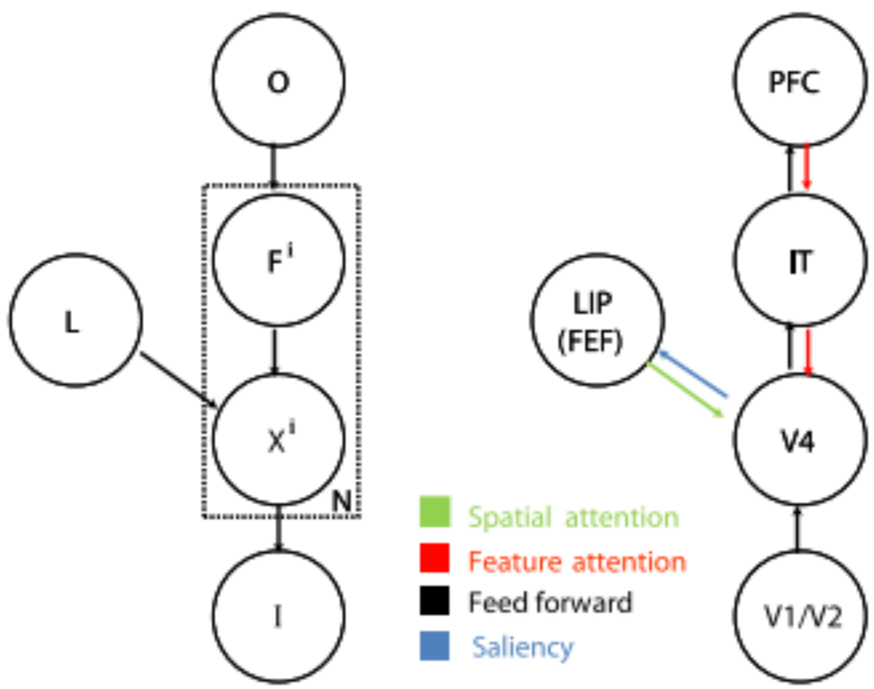


Fig. 2. Left: Proposed Bayesian model. Right: A model illustrating the interaction between the parietal and ventral streams mediated by feedforward and feedback connections. The main additions to the original feedforward model (Serre, Kouh, et al., 2005) (see also Supplementary Online Information) are (i) the cortical feedback within the ventral stream (providing feature-based attention); (ii) the cortical feedback from areas of the parietal cortex onto areas of the ventral stream (providing spatial attention); and (iii) feedforward connections to the parietal cortex that serves as a 'saliency map' encoding the visual relevance of image locations (Koch & Ullman, 1985).

N = number of objects

object encoding O (PFC)

feature encoding F (IT)

location encoding L (FEF)

joint location/feature map X^i (V4)

feedforward input (V1/V2)

Models

What and where: A Bayesian inference theory of attention

Sharat Chikkerur *, Thomas Serre, Cheston Tan, Tomaso Poggio

McGovern Institute for Brain Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, United States

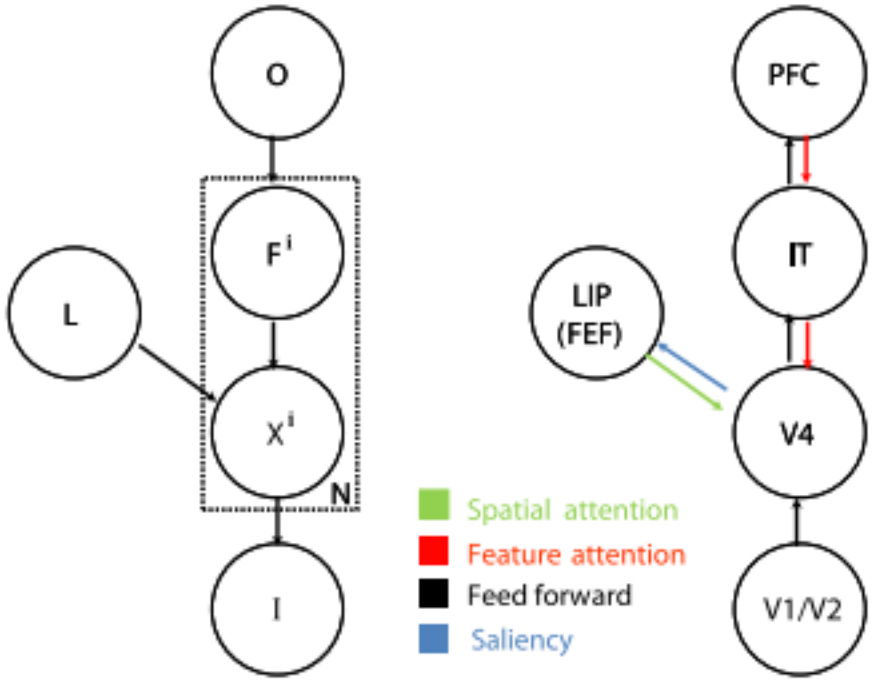


Fig. 2. Left: Proposed Bayesian model. Right: A model illustrating the interaction between the parietal and ventral streams mediated by feedforward and feedback connections. The main additions to the original feedforward model (Serre, Kouh, et al., 2005) (see also Supplementary Online Information) are (i) the cortical feedback within the ventral stream (providing feature-based attention); (ii) the cortical feedback from areas of the parietal cortex onto areas of the ventral stream (providing spatial attention); and (iii) feedforward connections to the parietal cortex that serves as a 'saliency map' encoding the visual relevance of image locations (Koch & Ullman, 1985).

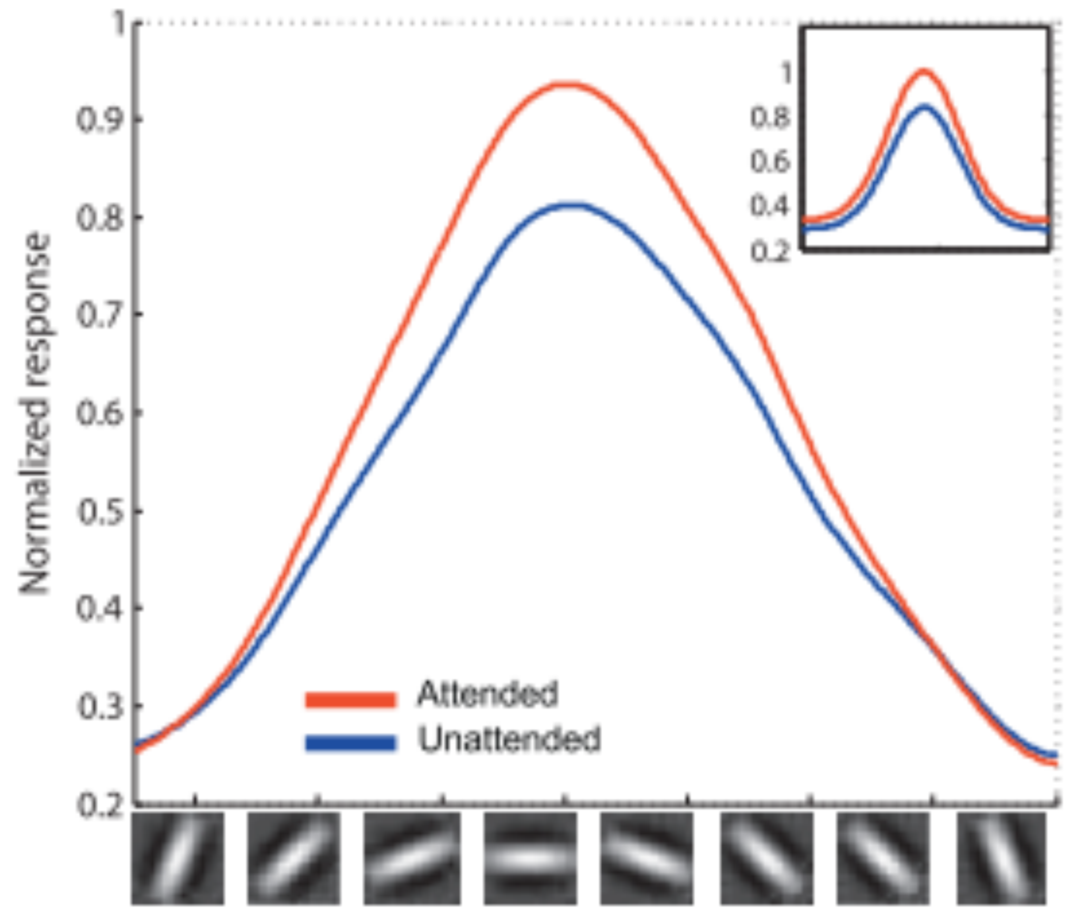
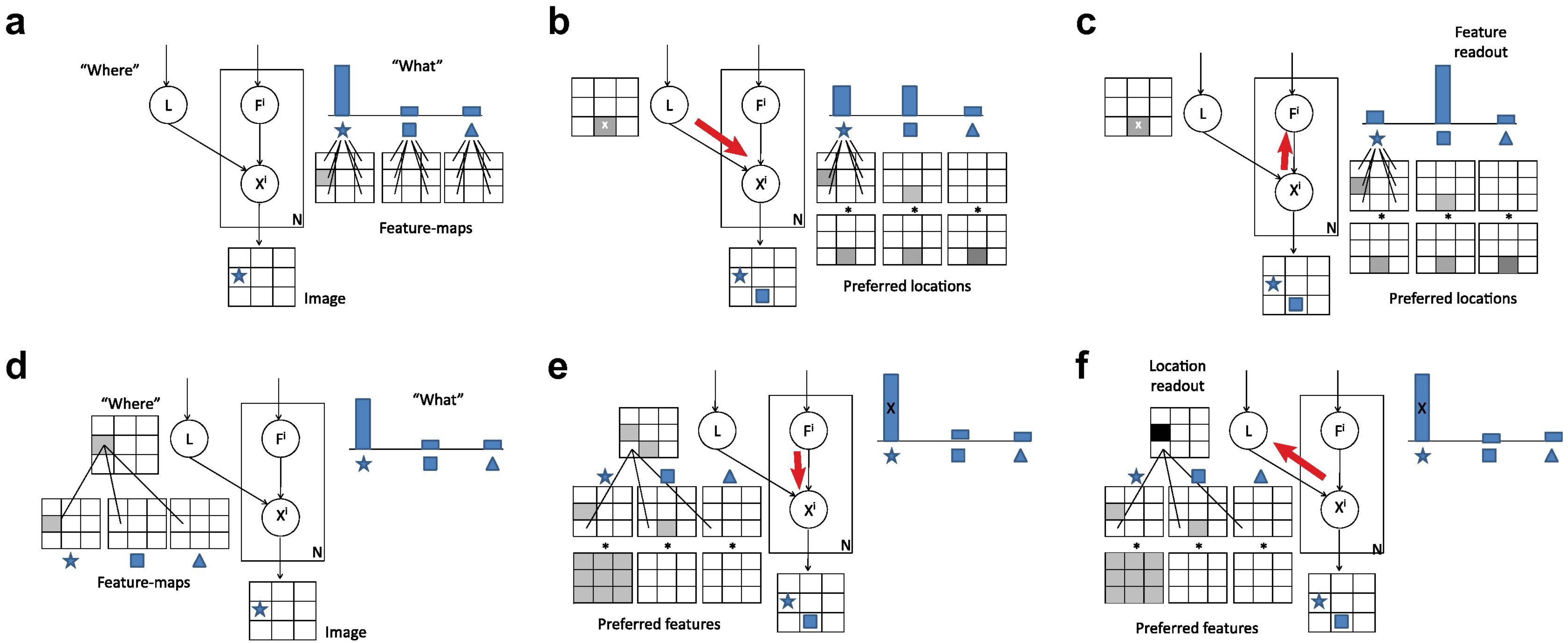


Fig. 4. Effect of spatial attention on tuning response. The tuning curve shows a multiplicative modulation under attention. The inset shows the replotted data from McAdams and Maunsell (1999).

Models

F^i = pools across locations in X^i



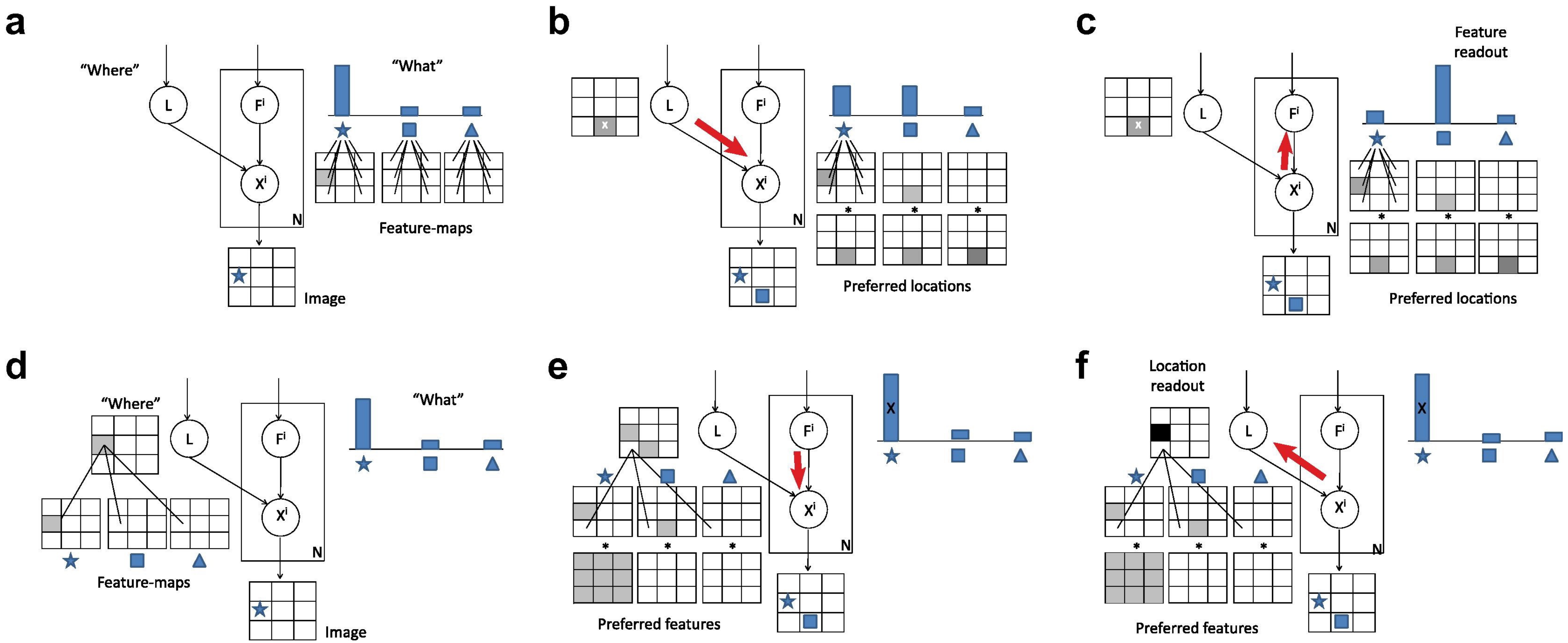
Models

What and where: A Bayesian inference theory of attention

Sharat Chikkerur *, Thomas Serre, Cheston Tan, Tomaso Poggio

McGovern Institute for Brain Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, United States

F^i = pools across locations in X^i   Spatial attention spotlight X



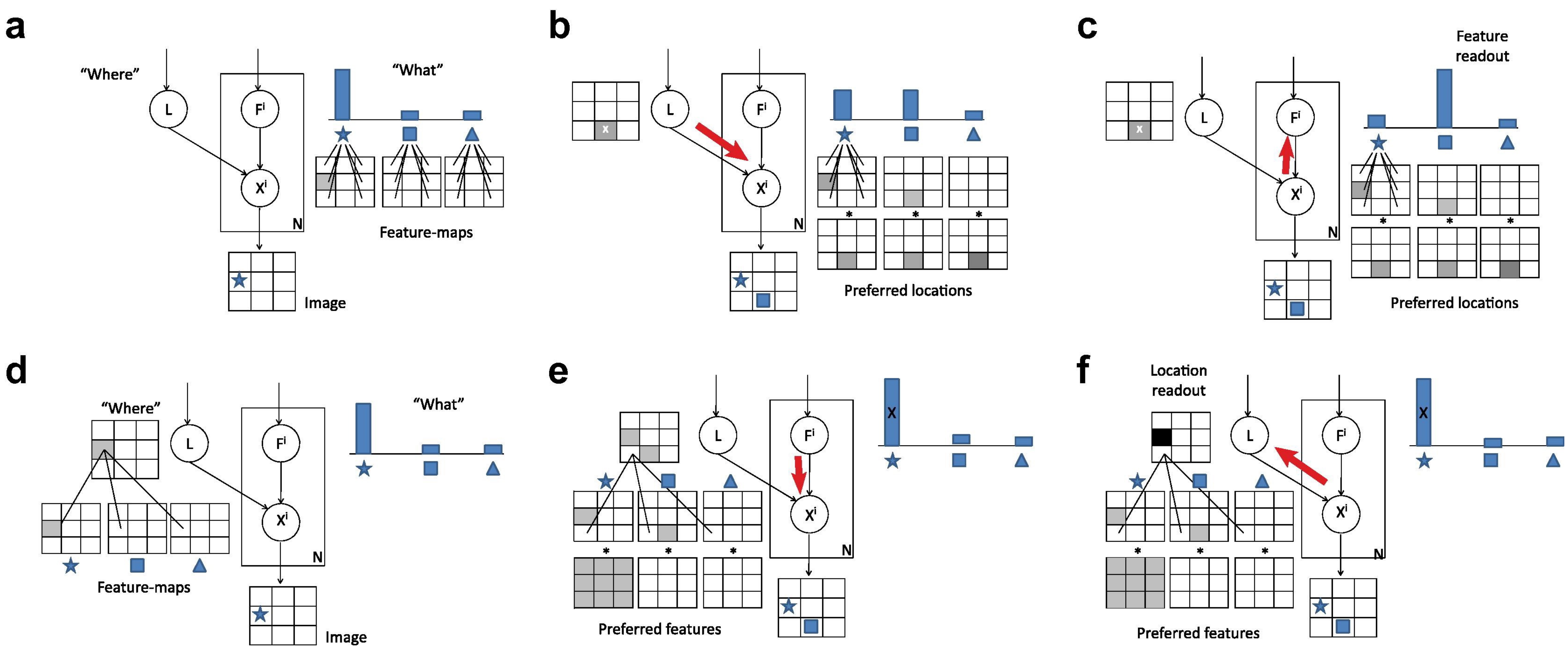
Models

What and where: A Bayesian inference theory of attention

Sharat Chikkerur *, Thomas Serre, Cheston Tan, Tomaso Poggio

McGovern Institute for Brain Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, United States

F^i = pools across locations in X^i   Spatial attention spotlight X   effect read out as P(F^i | I)



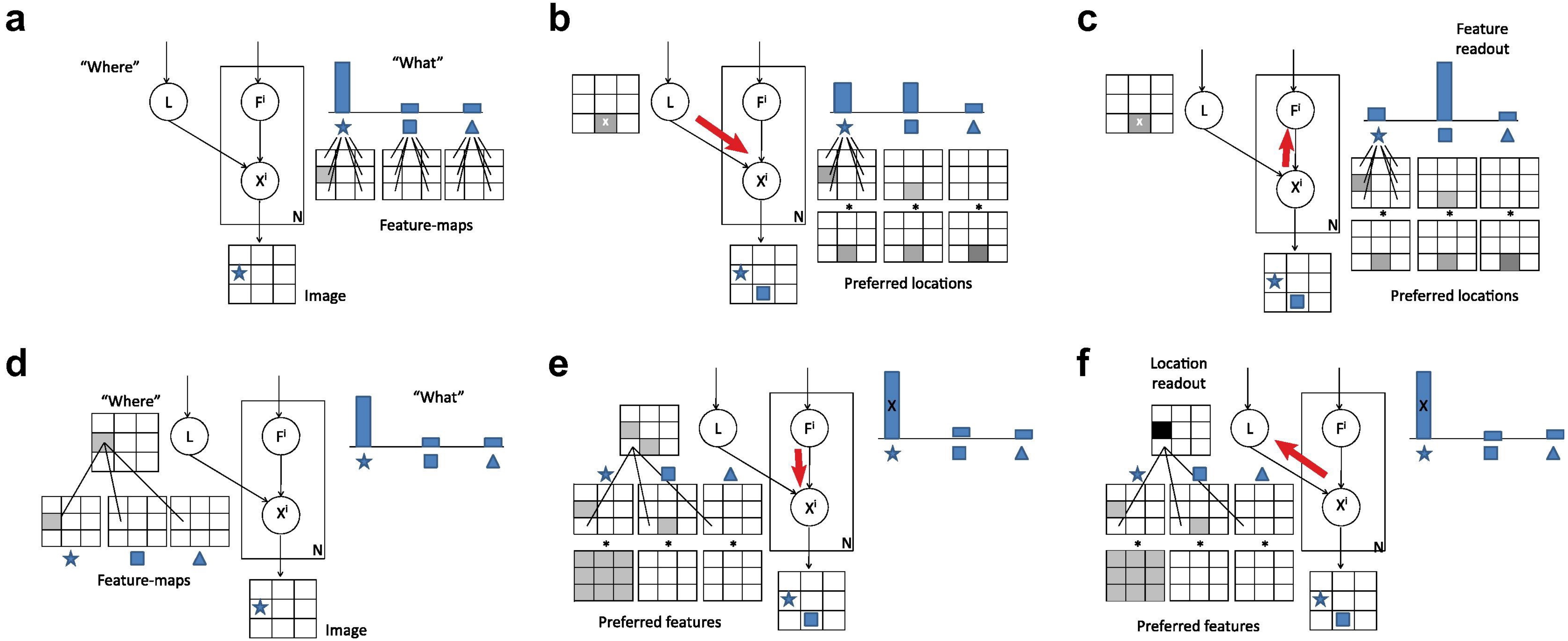
Models

What and where: A Bayesian inference theory of attention

Sharat Chikkerur *, Thomas Serre, Cheston Tan, Tomaso Poggio

McGovern Institute for Brain Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, United States

F^i = pools across locations in X^i   Spatial attention spotlight X   effect read out as P(F^i | I)



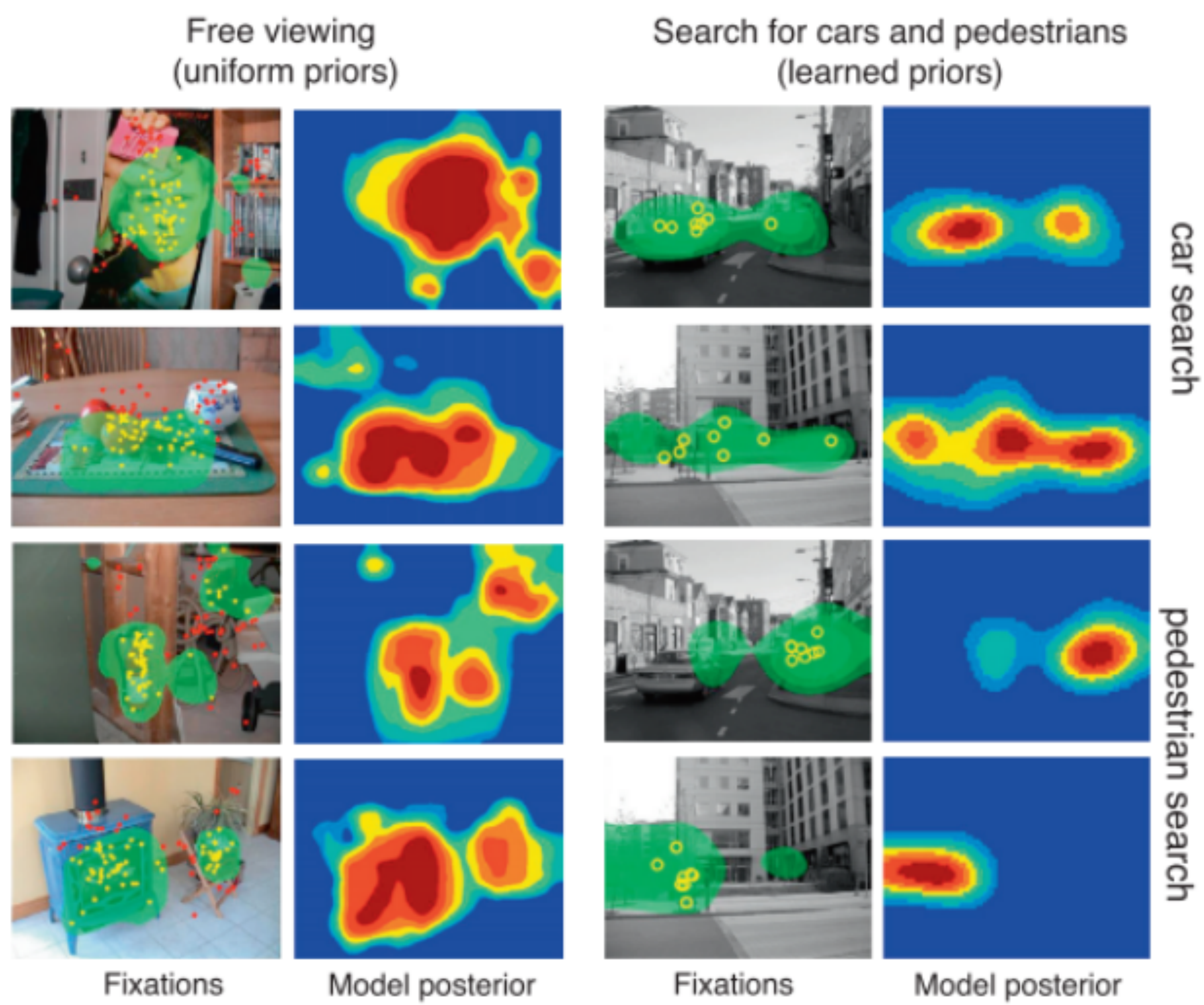
L represented in FEF   Feature attention makes P(F^i) high for preferred feature   Location of preferred feature read out as P(L | I)

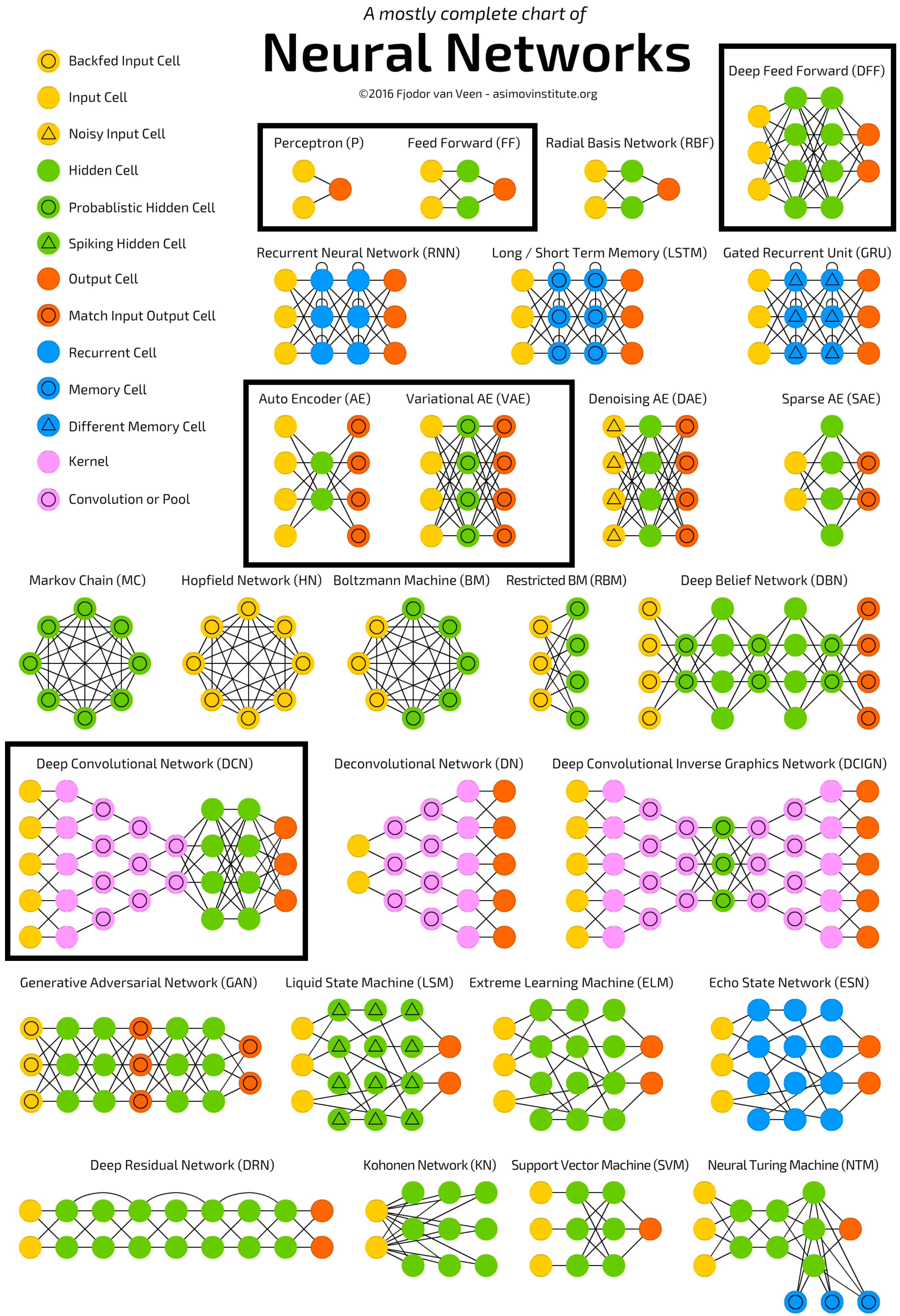
Models

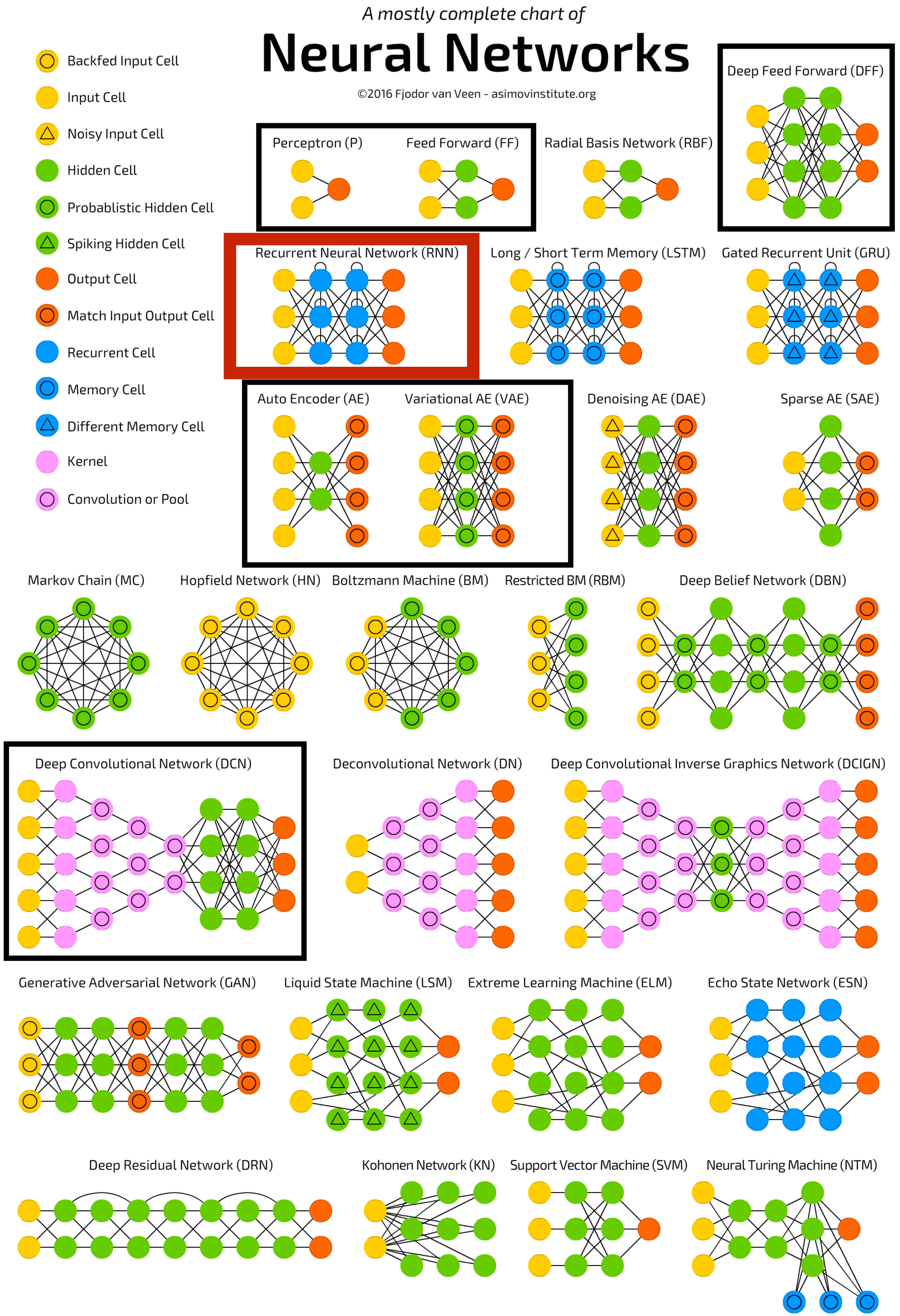
What and where: A Bayesian inference theory of attention

Sharat Chikkerur *, Thomas Serre, Cheston Tan, Tomaso Poggio

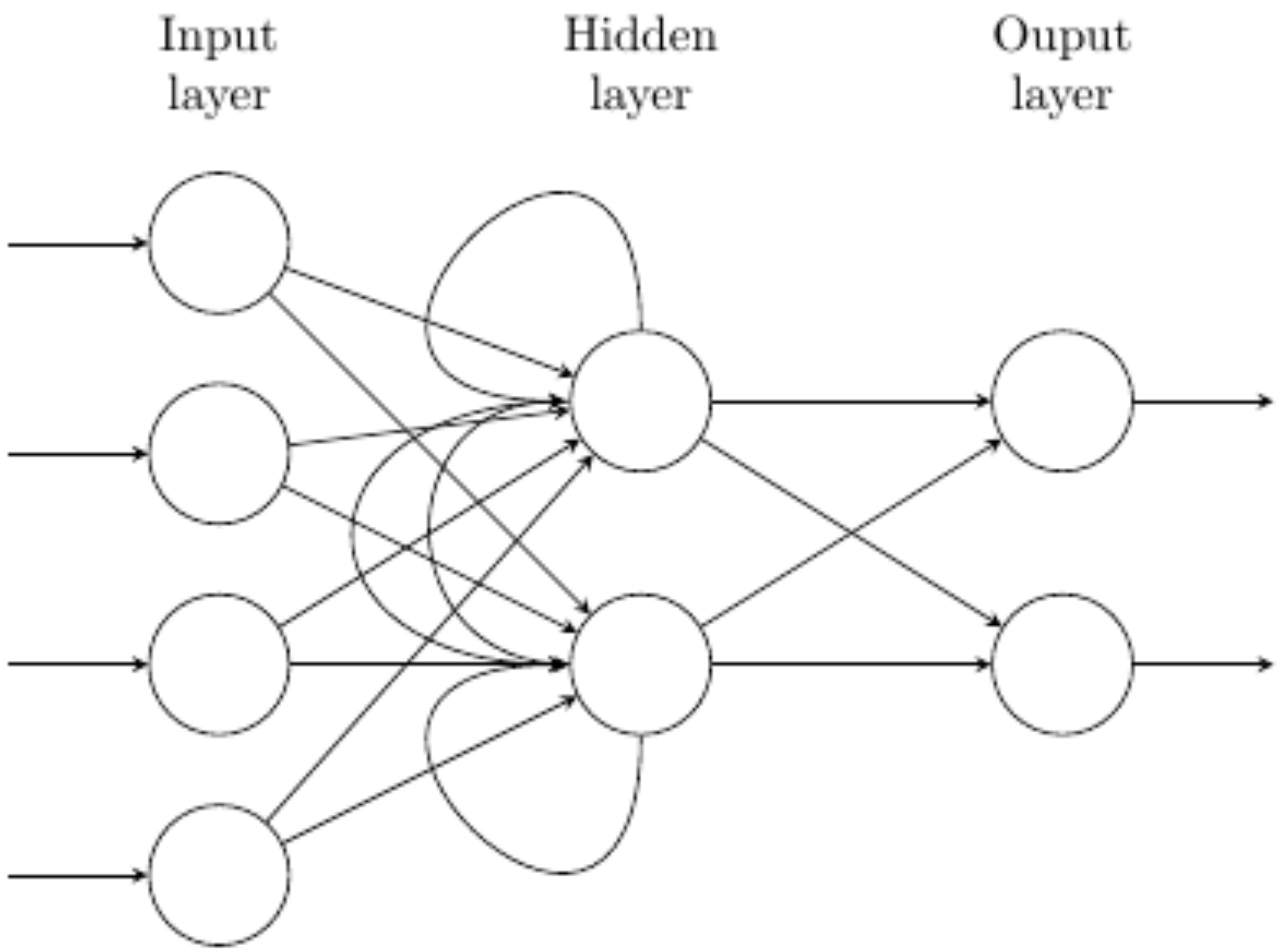
McGovern Institute for Brain Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, United States




A mostly complete chart of
Neural Networks
©2016 Fjodor van Veen - asimovinstitute.org
Backfed Input Cell
Input Cell
Noisy Input Cell
Hidden Cell
Probablistic Hidden Cell
Spiking Hidden Cell
Output Cell
Match Input Output Cell
Recurrent Cell
Memory Cell
Different Memory Cell
Kernel
Convolution or Pool
Perceptron (P)
Feed Forward (FF)
Radial Basis Network (RBF)
Deep Feed Forward (DFF)
Recurrent Neural Network (RNN)
Long / Short Term Memory (LSTM)
Gated Recurrent Unit (GRU)
Auto Encoder (AE)
Variational AE (VAE)
Denoising AE (DAE)
Sparse AE (SAE)
Markov Chain (MC)
Hopfield Network (HN)
Boltzmann Machine (BM)
Restricted BM (RBM)
Deep Belief Network (DBN)
Deep Convolutional Network (DCN)
Deconvolutional Network (DN)
Deep Convolutional Inverse Graphics Network (DCIGN)
Generative Adversarial Network (GAN)
Liquid State Machine (LSM)
Extreme Learning Machine (ELM)
Echo State Network (ESN)
Deep Residual Network (DRN)
Kohonen Network (KN)
Support Vector Machine (SVM)
Neural Turing Machine (NTM)

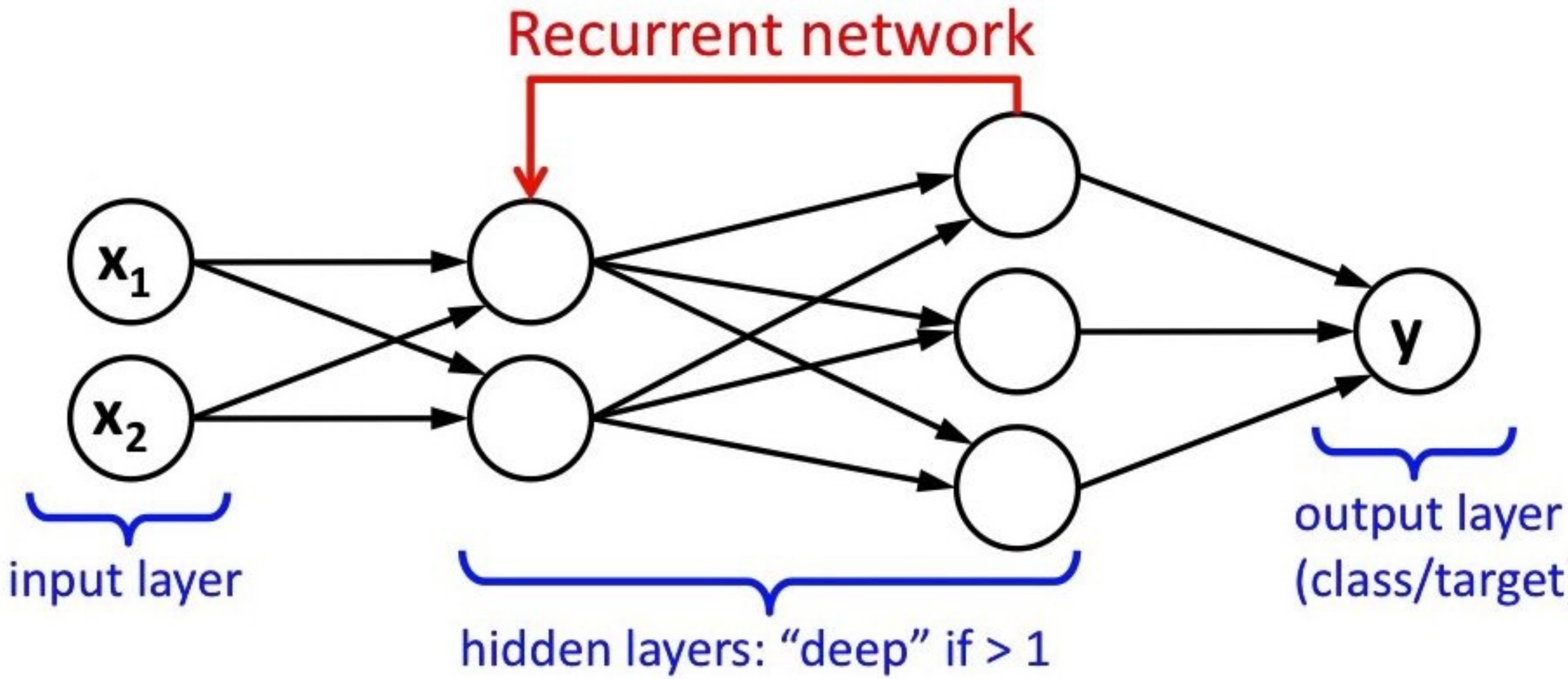

A mostly complete chart of
Neural Networks
©2016 Fjodor van Veen - asimovinstitute.org
Backfed Input Cell
Input Cell
Noisy Input Cell
Hidden Cell
Probablistic Hidden Cell
Spiking Hidden Cell
Output Cell
Match Input Output Cell
Recurrent Cell
Memory Cell
Different Memory Cell
Kernel
Convolution or Pool
Perceptron (P)
Feed Forward (FF)
Radial Basis Network (RBF)
Deep Feed Forward (DFF)
Recurrent Neural Network (RNN)
Long / Short Term Memory (LSTM)
Gated Recurrent Unit (GRU)
Auto Encoder (AE)
Variational AE (VAE)
Denoising AE (DAE)
Sparse AE (SAE)
Markov Chain (MC)
Hopfield Network (HN)
Boltzmann Machine (BM)
Restricted BM (RBM)
Deep Belief Network (DBN)
Deep Convolutional Network (DCN)
Deconvolutional Network (DN)
Deep Convolutional Inverse Graphics Network (DCIGN)
Generative Adversarial Network (GAN)
Liquid State Machine (LSM)
Extreme Learning Machine (ELM)
Echo State Network (ESN)
Deep Residual Network (DRN)
Kohonen Network (KN)
Support Vector Machine (SVM)
Neural Turing Machine (NTM)

Simple (unrestricted) RNNs

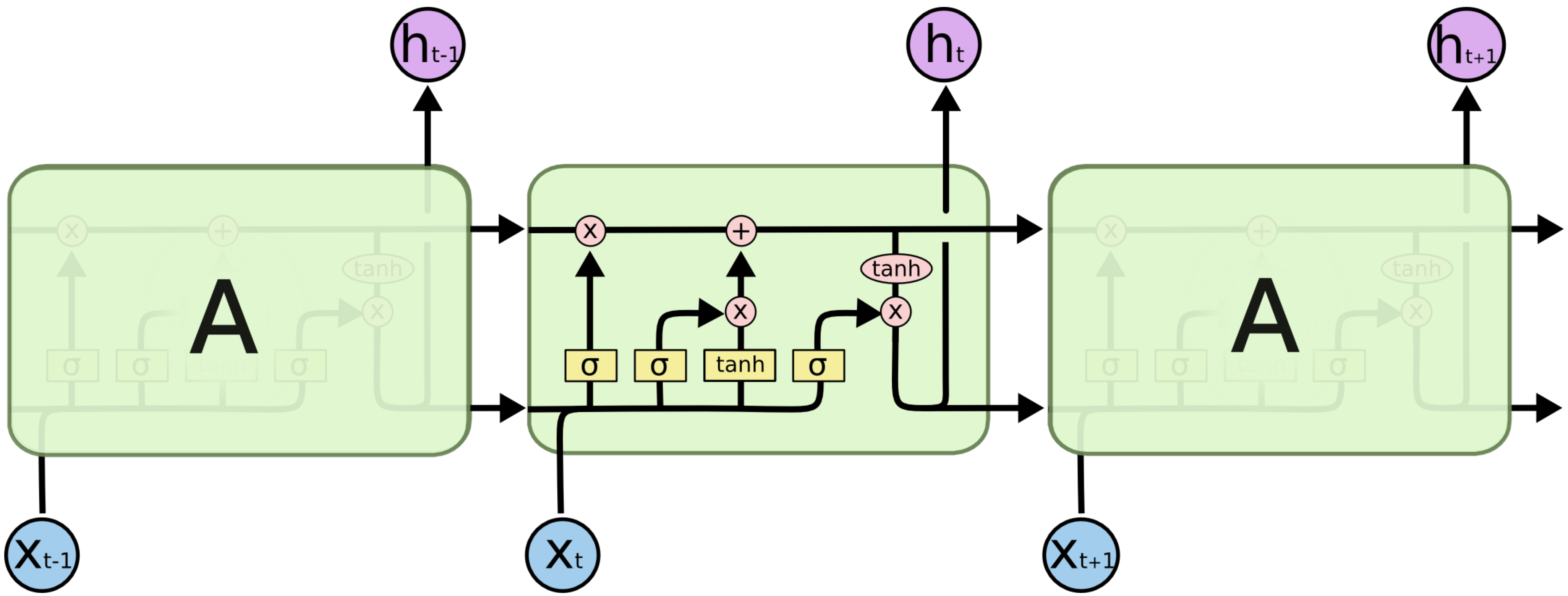


Unlike feedforward networks, recurrent networks can **store state**.

Simple (unrestricted) RNNs

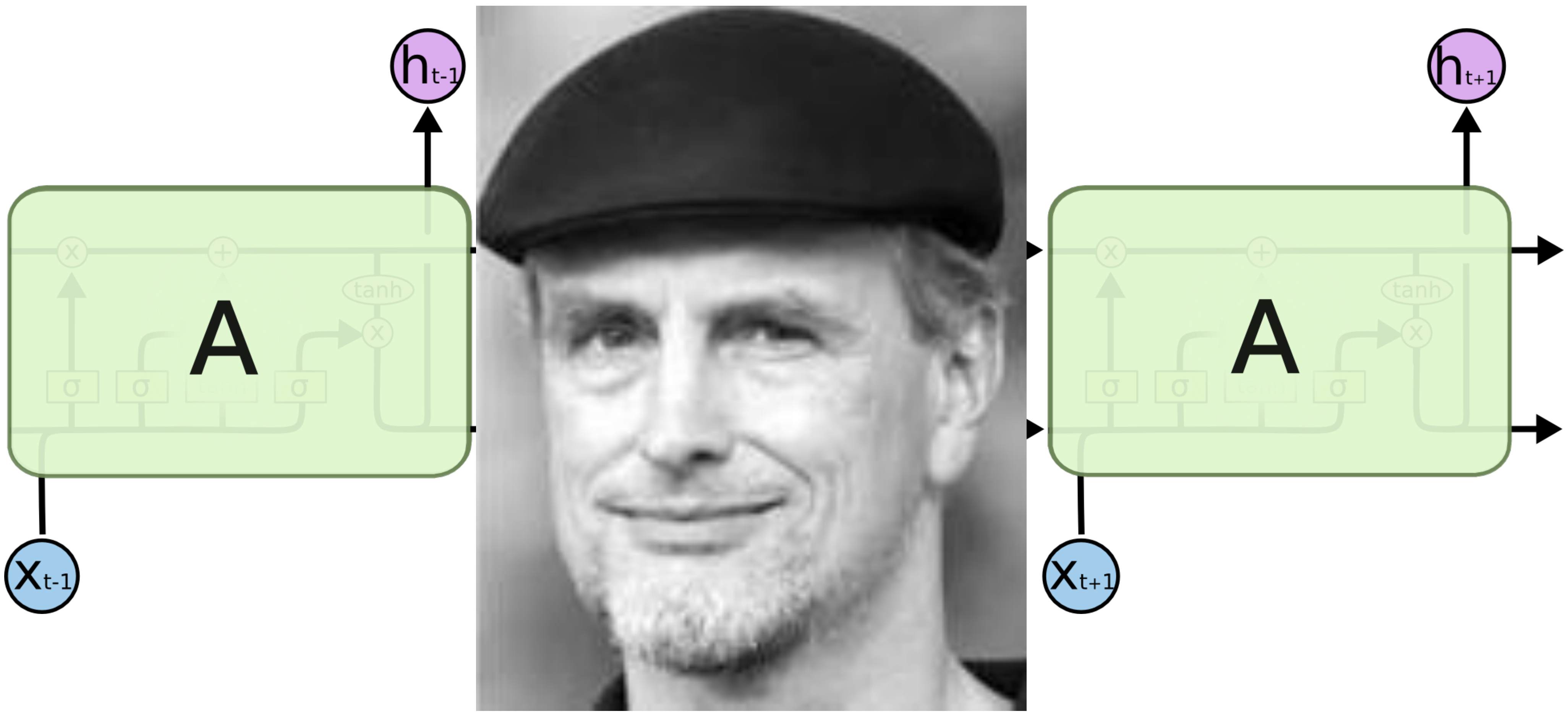


Long-Short Term Memory (LSTMs)



Models: RNNs

Long-Short Term Memory (LSTMs)

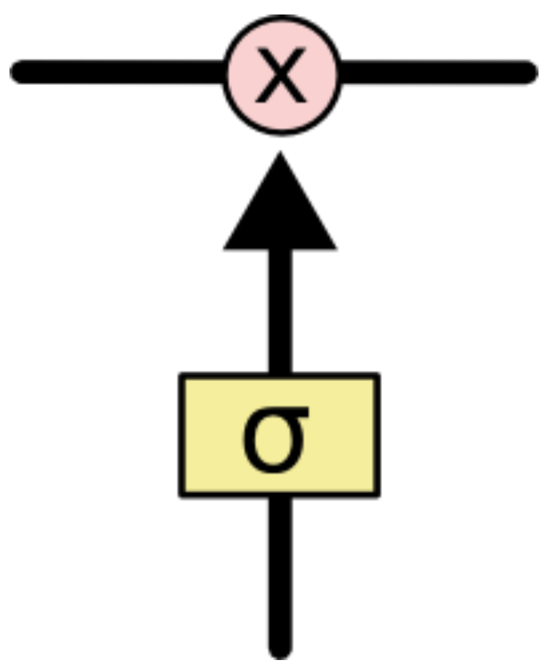


Jürgen Schmidhuber

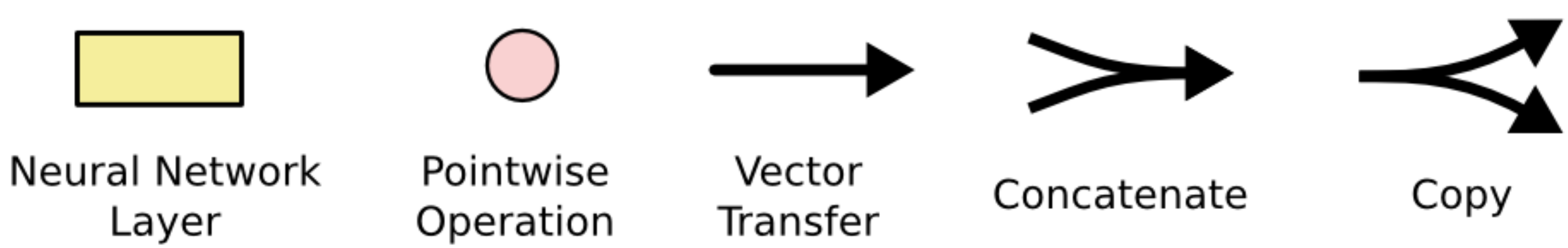
Models: RNNs

Long-Short Term Memory (LSTMs)

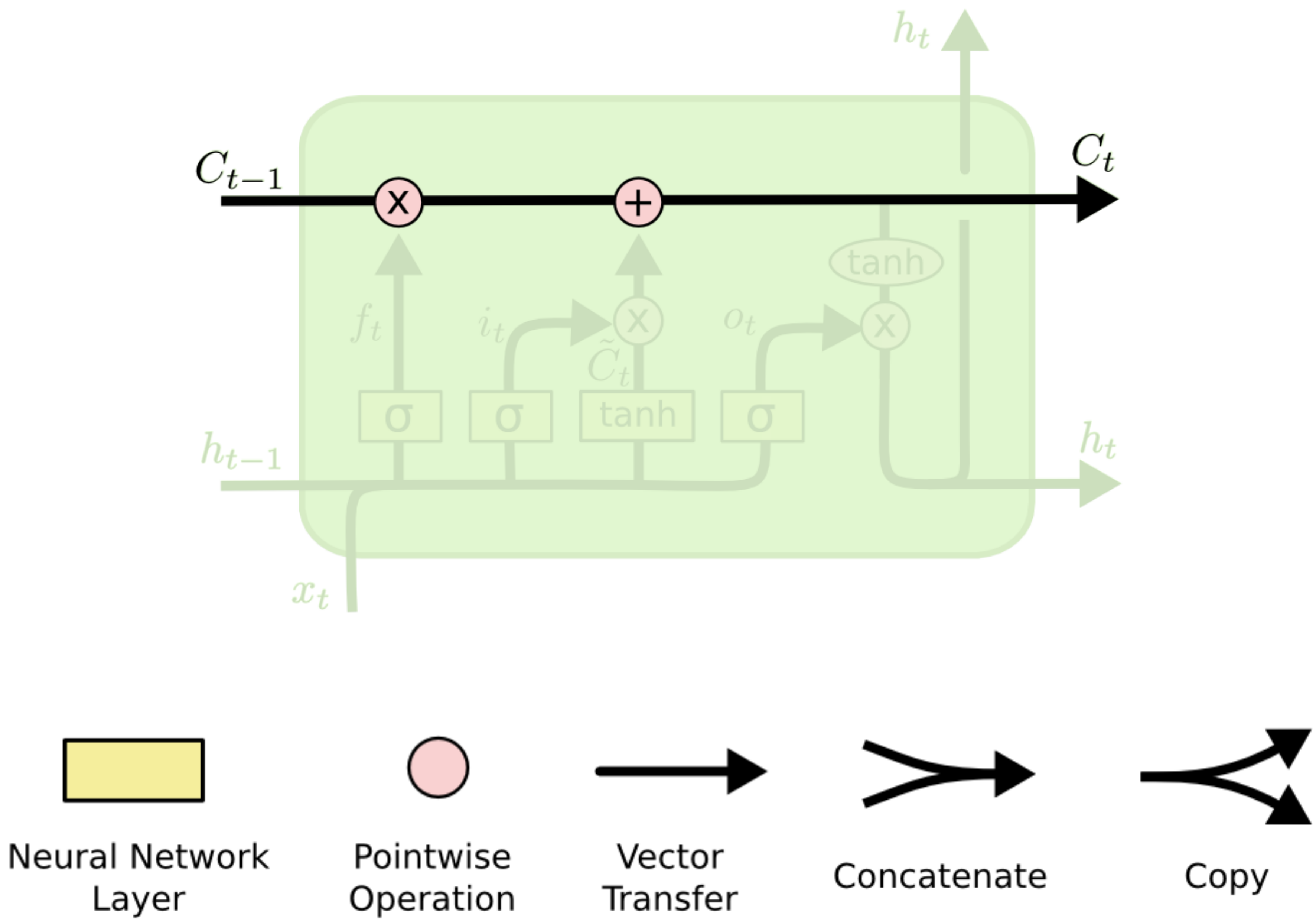
Gate:



sigmoid + pointwise multiplication

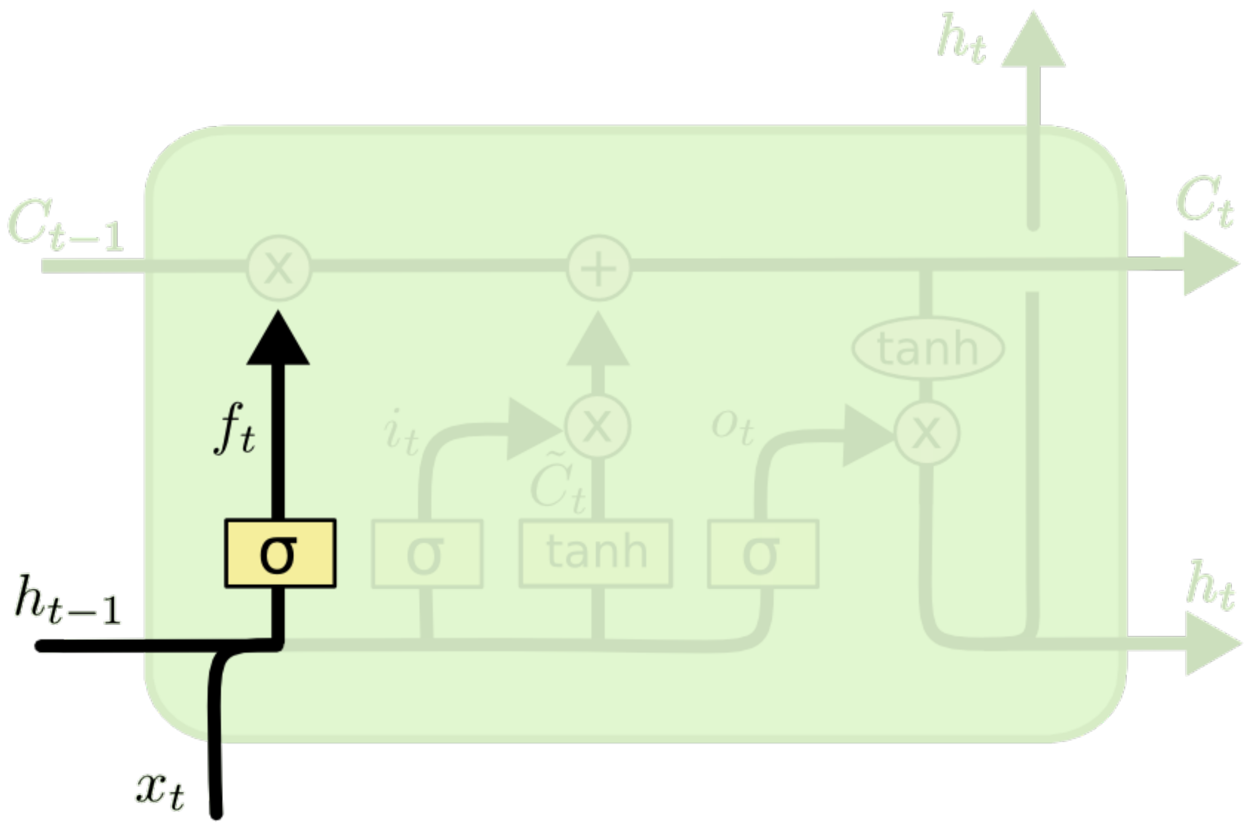


Long-Short Term Memory (LSTMs)



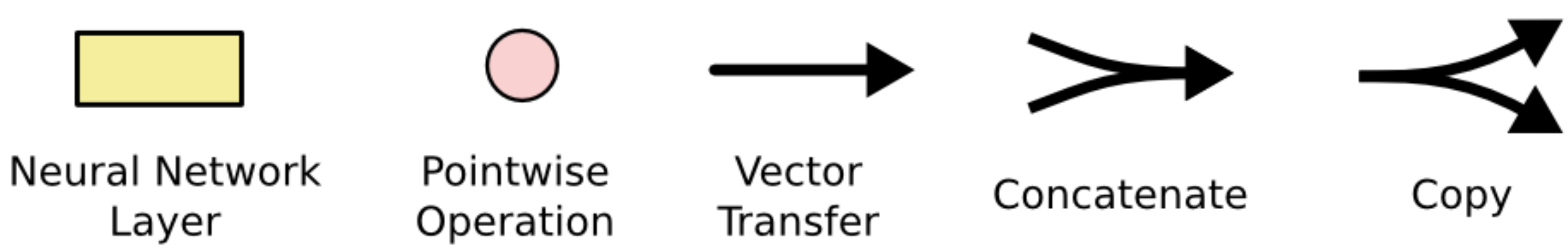
Models: RNNs

Long-Short Term Memory (LSTMs)



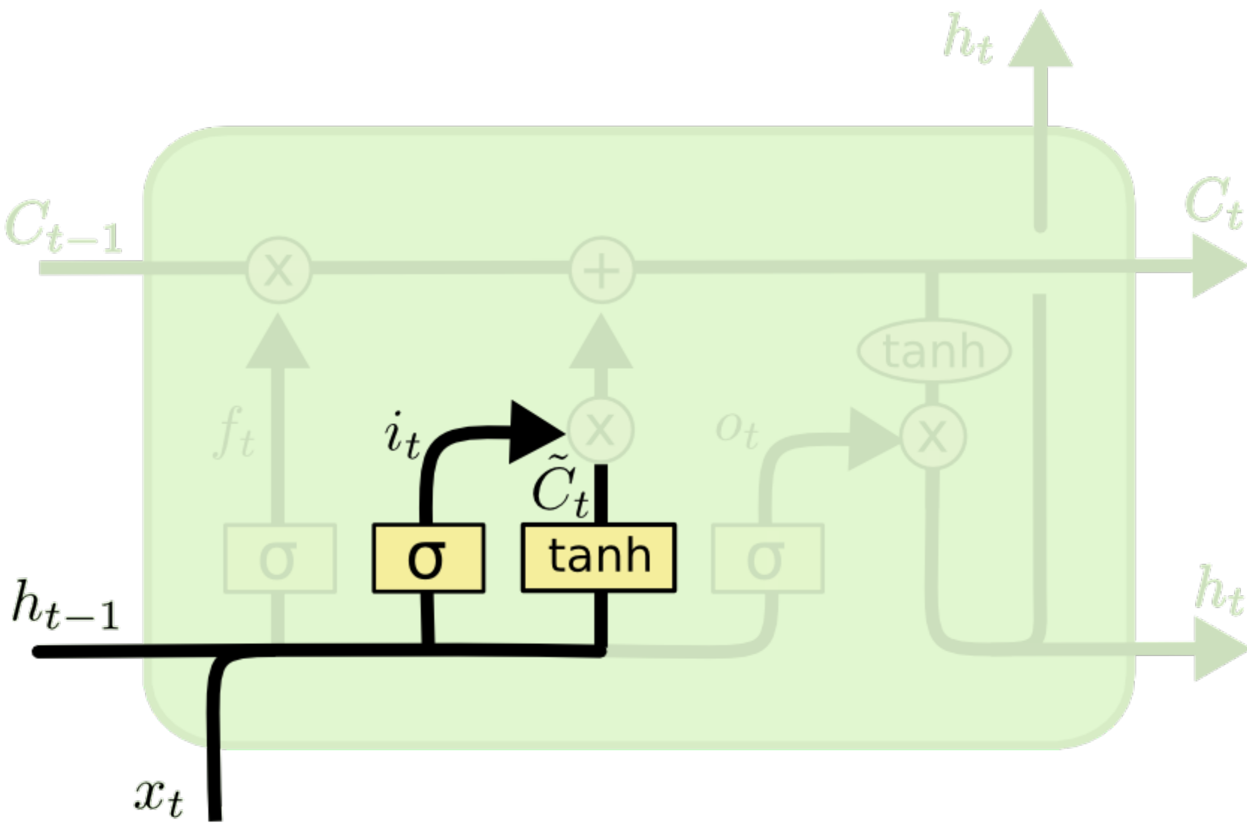
what to throw away:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$



Models: RNNs

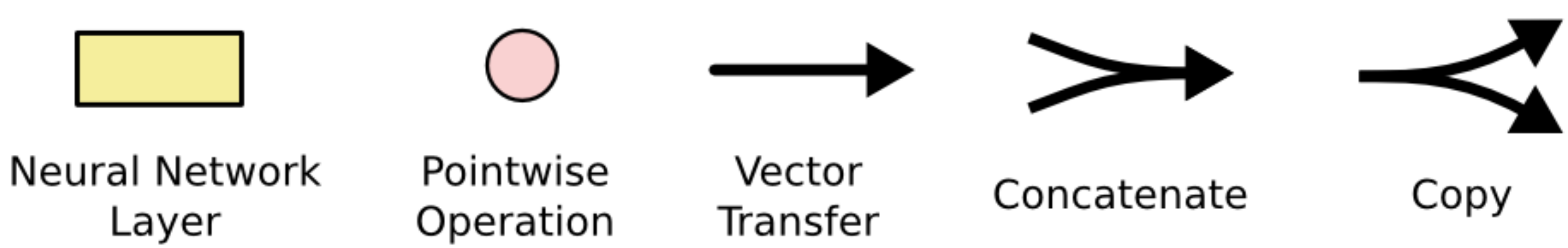
Long-Short Term Memory (LSTMs)



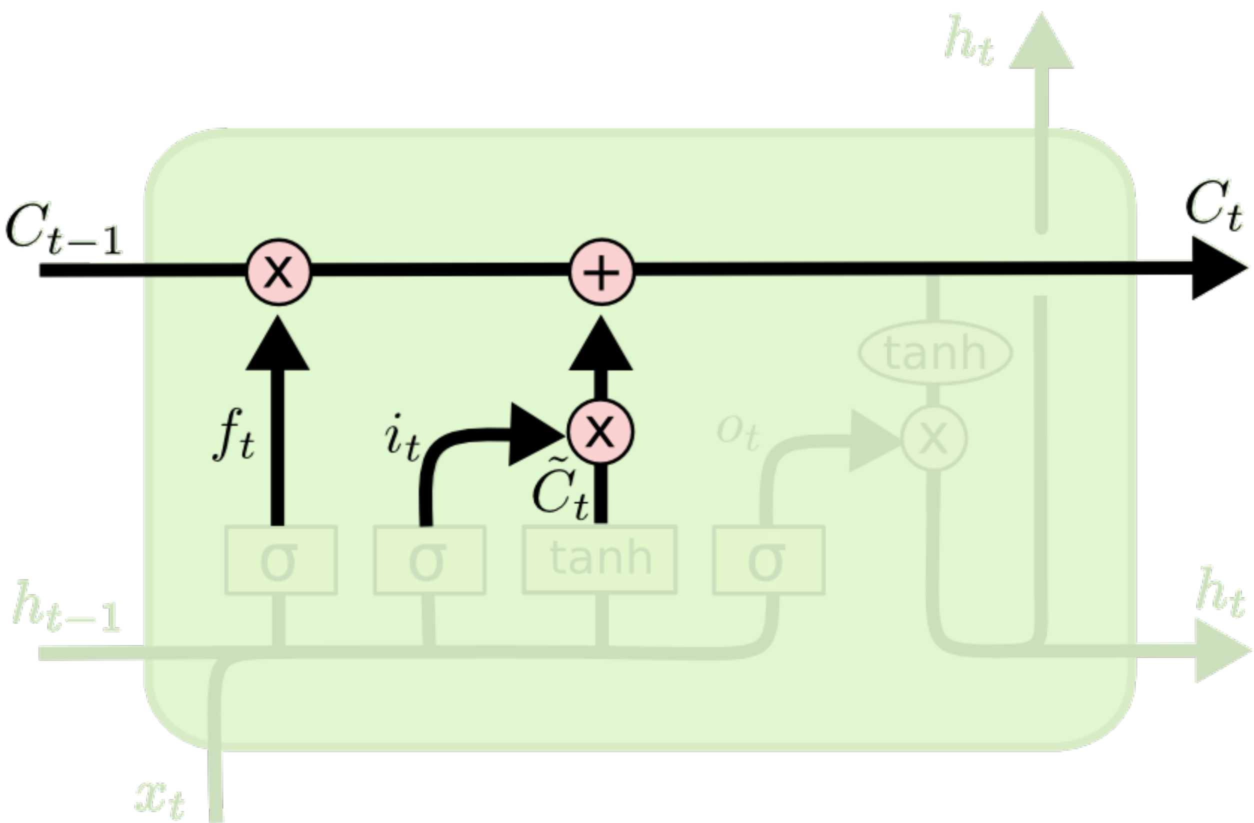
what to store:

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \ + \ b_C)$$

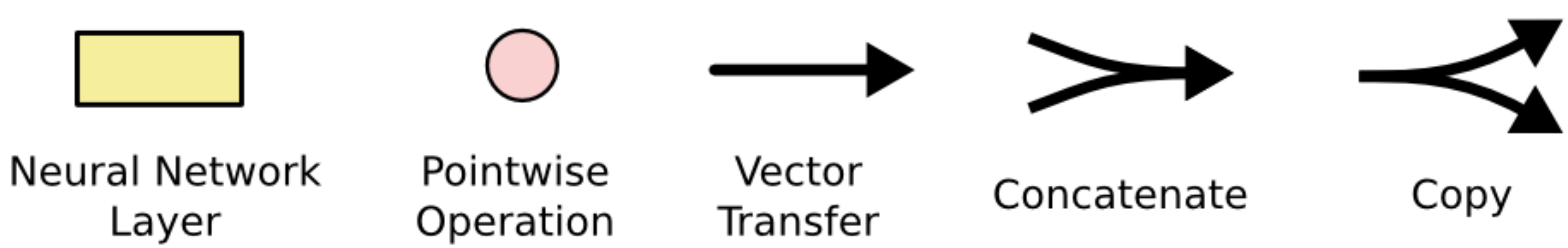


Long-Short Term Memory (LSTMs)

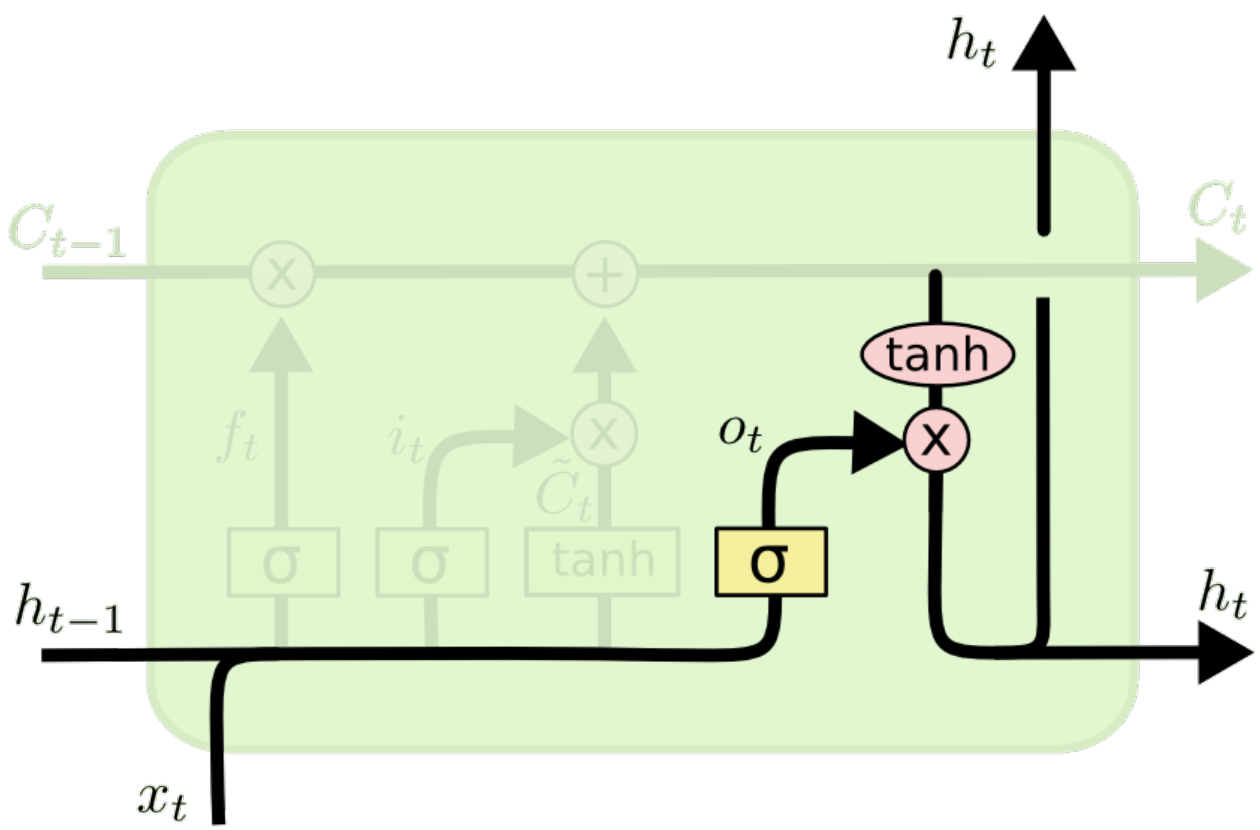


update old cell state:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$



Long-Short Term Memory (LSTMs)

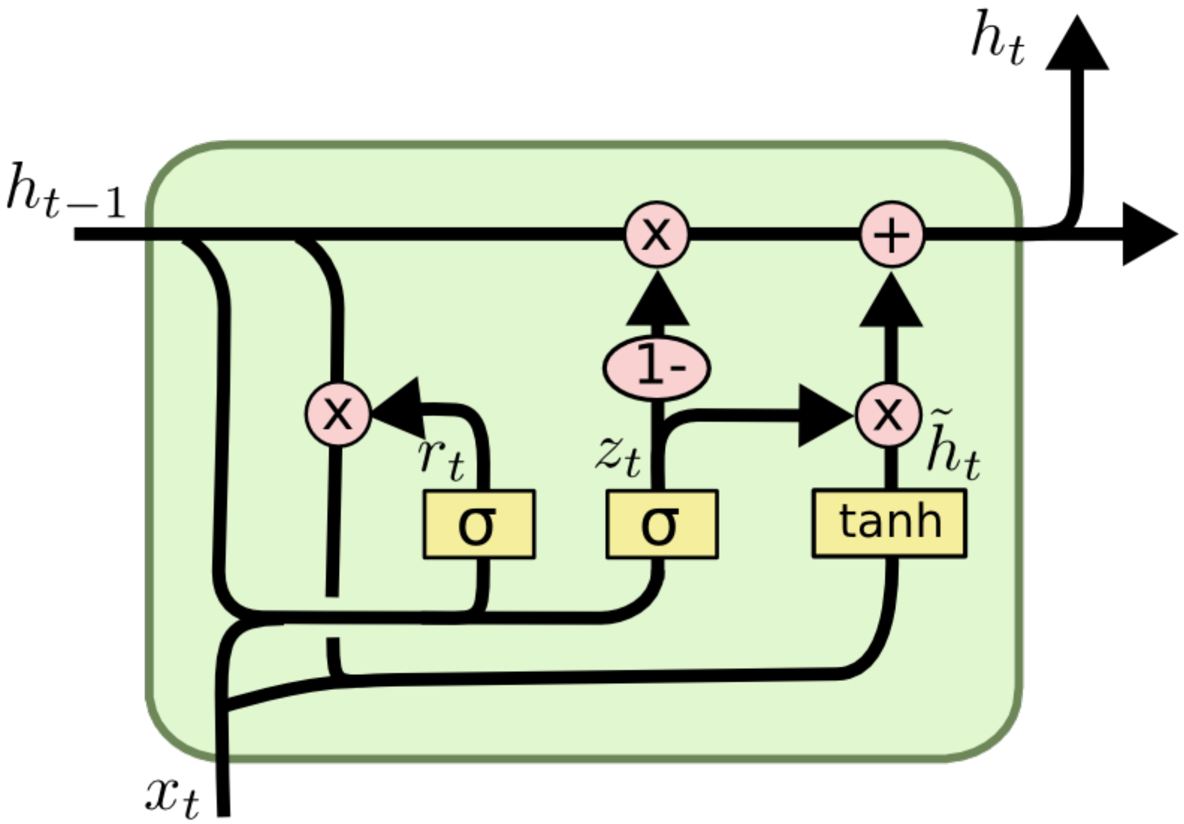


what to actually output:

$$o_t = \sigma\left(W_o\ [h_{t-1}, x_t]\ +\ b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

Gated Recurrent Unit (GRU) combines forget and input gate:

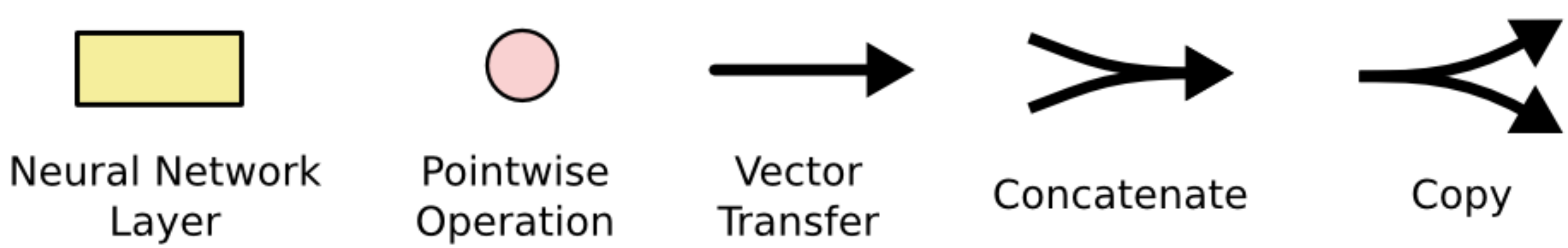


$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$



Models

Recurrent convolutional neural networks suppress occluders and enhance targets in occluded object recognition

Courtney J. Spoerer (courtney.spoerer@mrc-cbu.cam.ac.uk)
Medical Research Council Cognition and Brain Sciences Unit,
15 Chaucer Road, Cambridge, CB2 7EF, UK

Nikolaus Kriegeskorte (nikokriegeskorte@gmail.com)
Medical Research Council Cognition and Brain Sciences Unit,
15 Chaucer Road, Cambridge, CB2 7EF, UK

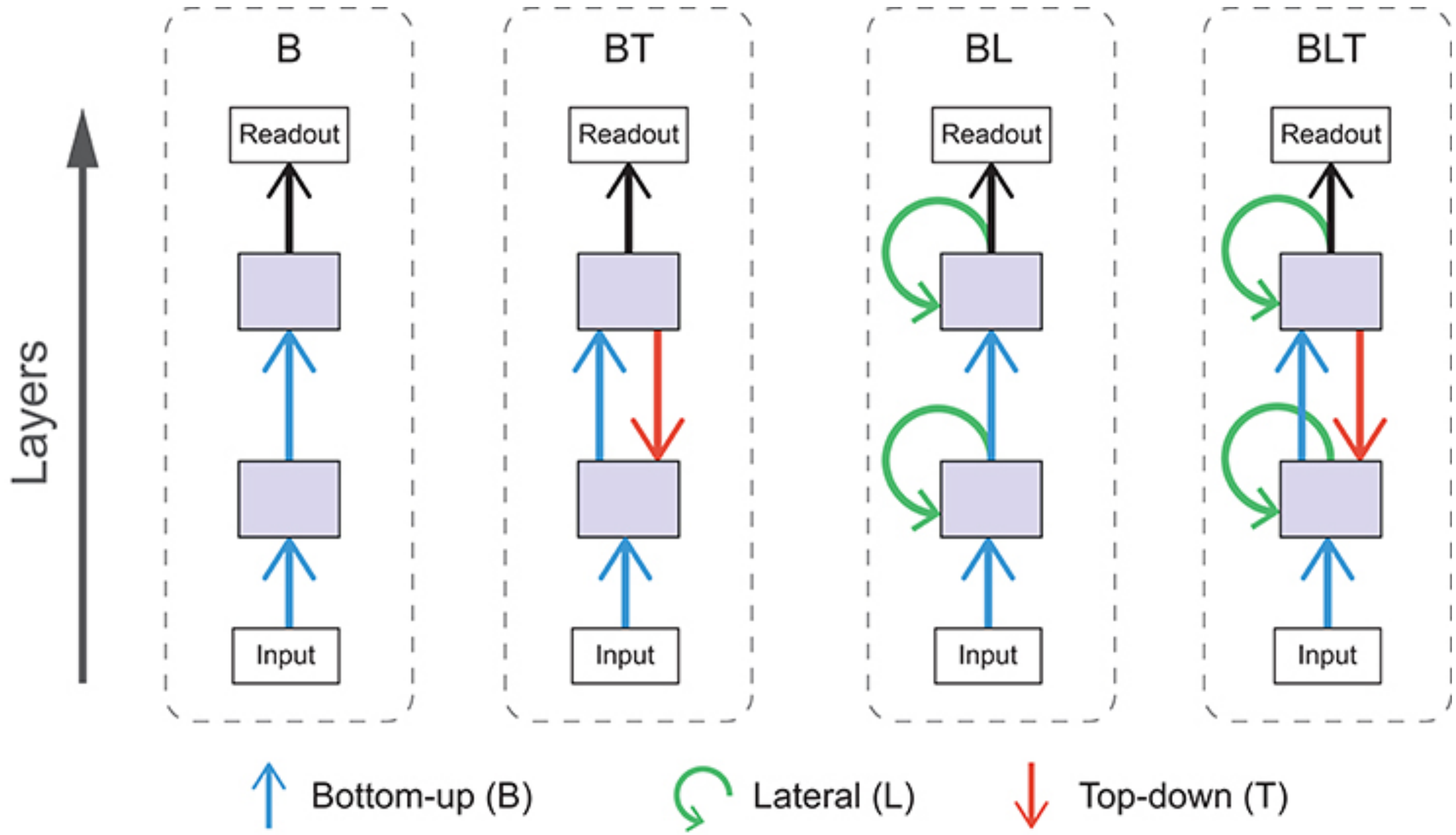
Figure 1: The process for generating stimuli for digit debris. First the target digit is generated. Random crops of all possible targets are taken to create a mask of debris, which is applied to the target as an occluder.

Models

Recurrent convolutional neural networks suppress occluders and enhance targets in occluded object recognition

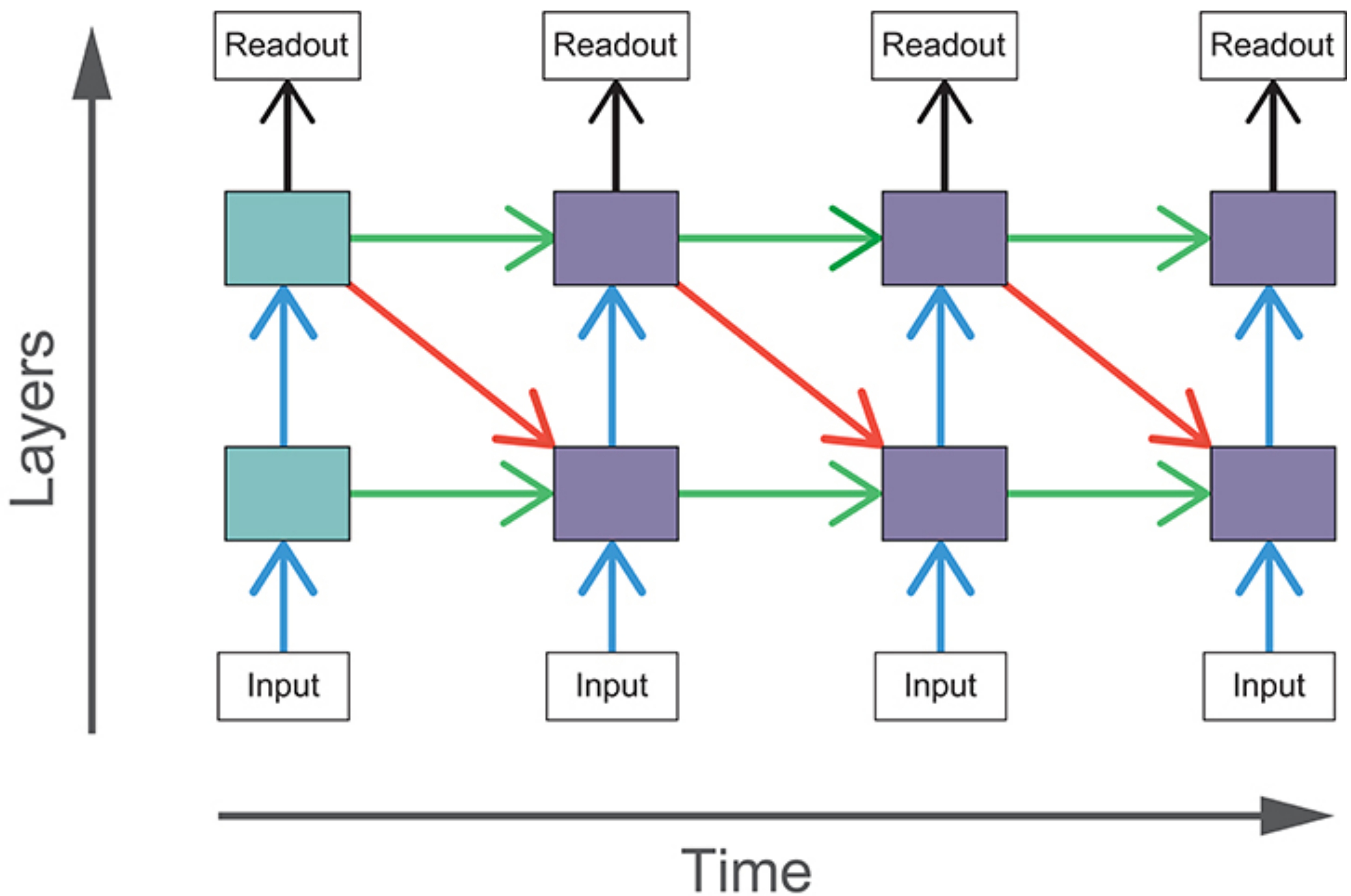
Courtney J. Spoerer (courtney.spoerer@mrc-cbu.cam.ac.uk)
Medical Research Council Cognition and Brain Sciences Unit,
15 Chaucer Road, Cambridge, CB2 7EF, UK

Nikolaus Kriegeskorte (nikokriegeskorte@gmail.com)
Medical Research Council Cognition and Brain Sciences Unit,
15 Chaucer Road, Cambridge, CB2 7EF, UK



Models

All RNNs executed by <u>unrolling in time</u>



Models

Recurrent convolutional neural networks suppress occluders and enhance targets in occluded object recognition

Courtney J. Spoerer (courtney.spoerer@mrc-cbu.cam.ac.uk)
Medical Research Council Cognition and Brain Sciences Unit,
15 Chaucer Road, Cambridge, CB2 7EF, UK

Nikolaus Kriegeskorte (nikokriegeskorte@gmail.com)
Medical Research Council Cognition and Brain Sciences Unit,
15 Chaucer Road, Cambridge, CB2 7EF, UK

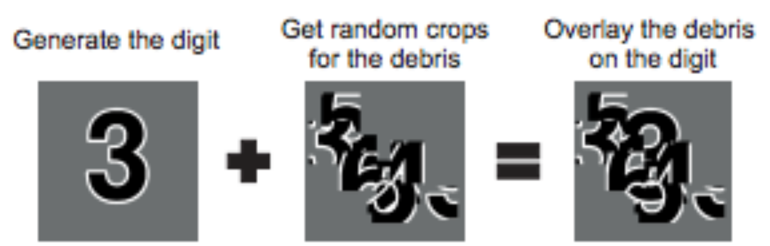


Figure 1: The process for generating stimuli for digit debris. First the target digit is generated. Random crops of all possible targets are taken to create a mask of debris, which is applied to the target as an occluder.

Results

Recurrent networks significantly out-performed feedforward networks across varying levels of occlusion. The difference in performance between feedforward and recurrent networks increased as the occlusion increased (Figure 2).

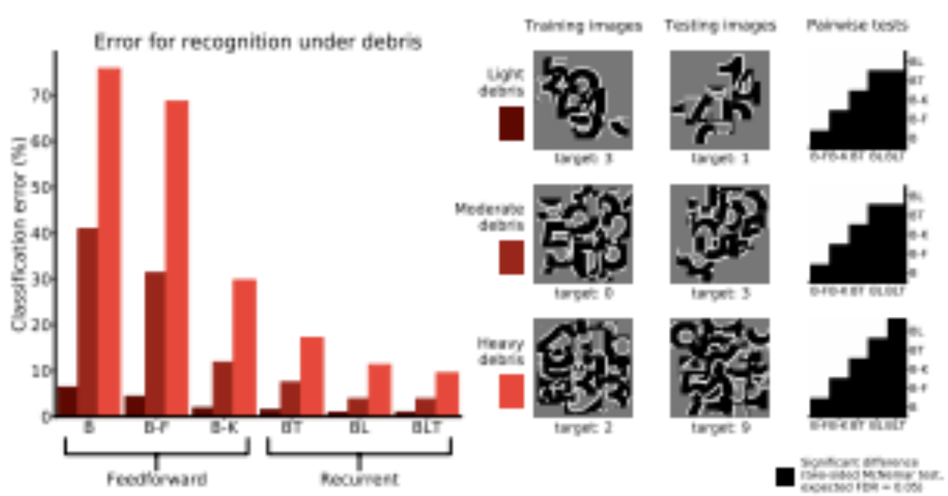


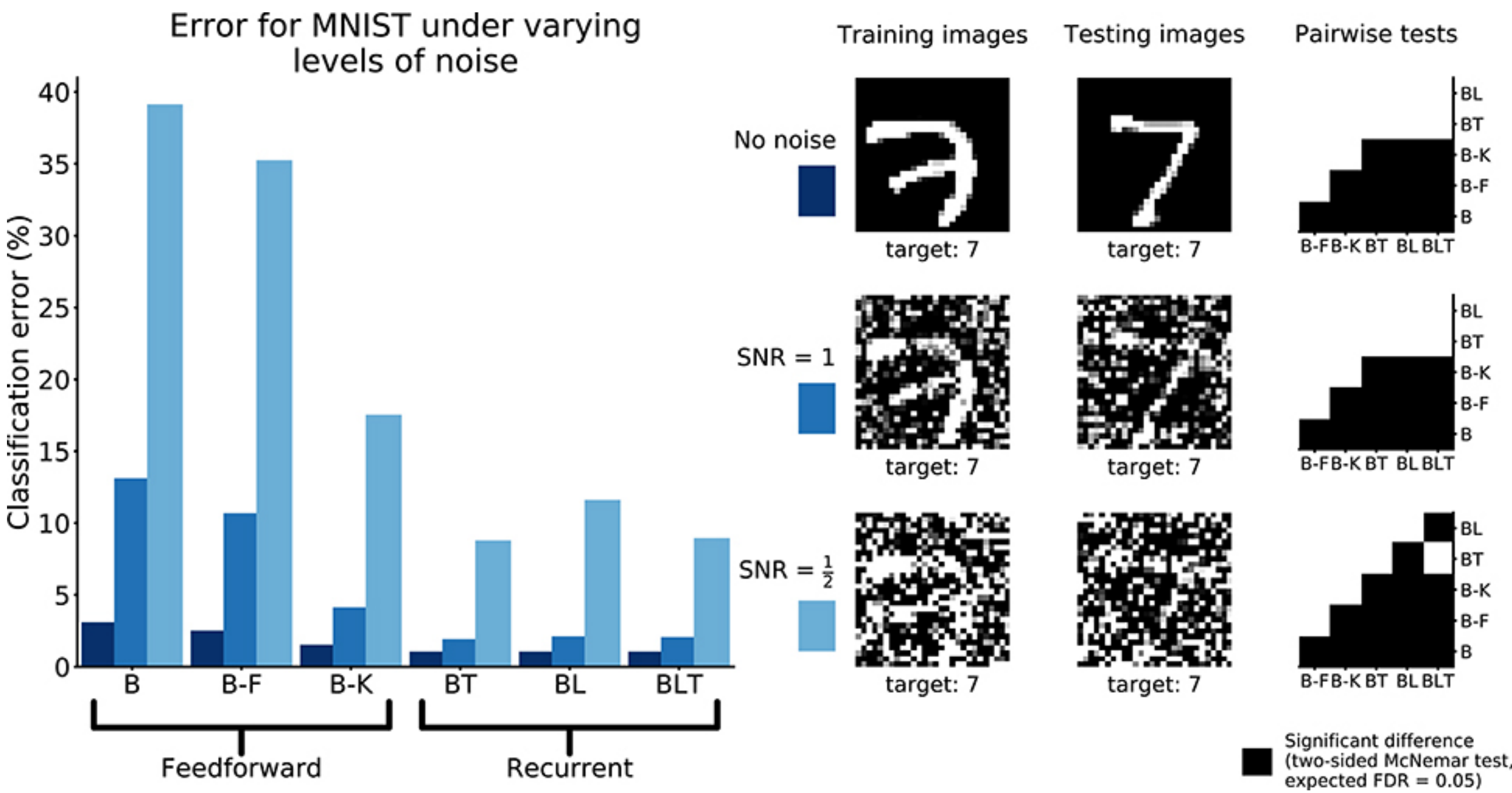
Figure 2: Classification error of the networks across increasing levels of debris (left). Pairwise differences across architectures for different levels of debris are indicated in matrix form (right).

Models

Recurrent convolutional neural networks suppress occluders and enhance targets in occluded object recognition

Courtney J. Spoerer (courtney.spoerer@mrc-cbu.cam.ac.uk)
Medical Research Council Cognition and Brain Sciences Unit,
15 Chaucer Road, Cambridge, CB2 7EF, UK

Nikolaus Kriegeskorte (nikokriegeskorte@gmail.com)
Medical Research Council Cognition and Brain Sciences Unit,
15 Chaucer Road, Cambridge, CB2 7EF, UK



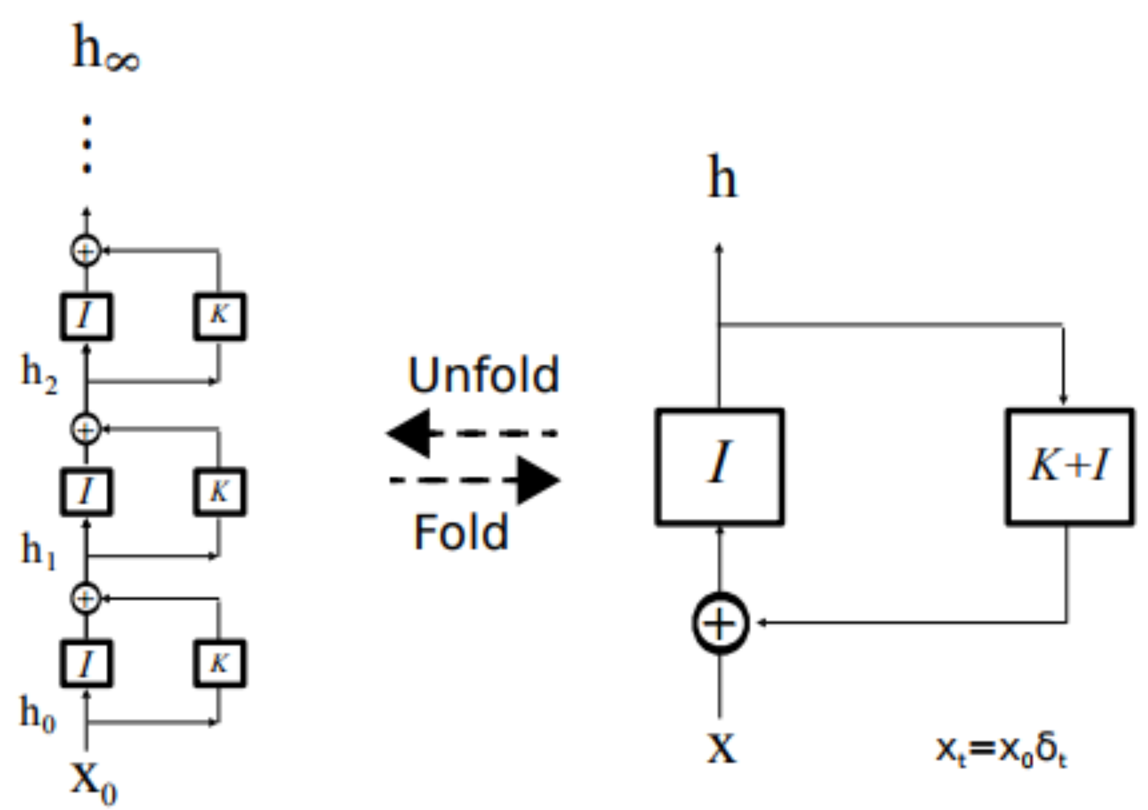
Models

CBMM Memo No. 047 April 12, 2016

Bridging the Gaps Between Residual Learning, Recurrent Neural Networks and Visual Cortex

by

Qianli Liao and Tomaso Poggio
Center for Brains, Minds and Machines, McGovern Institute, MIT



(A) ResNet with shared weights (B) ResNet in recurrent form

Models

Bridging the Gaps Between Residual Learning, Recurrent Neural Networks and Visual Cortex

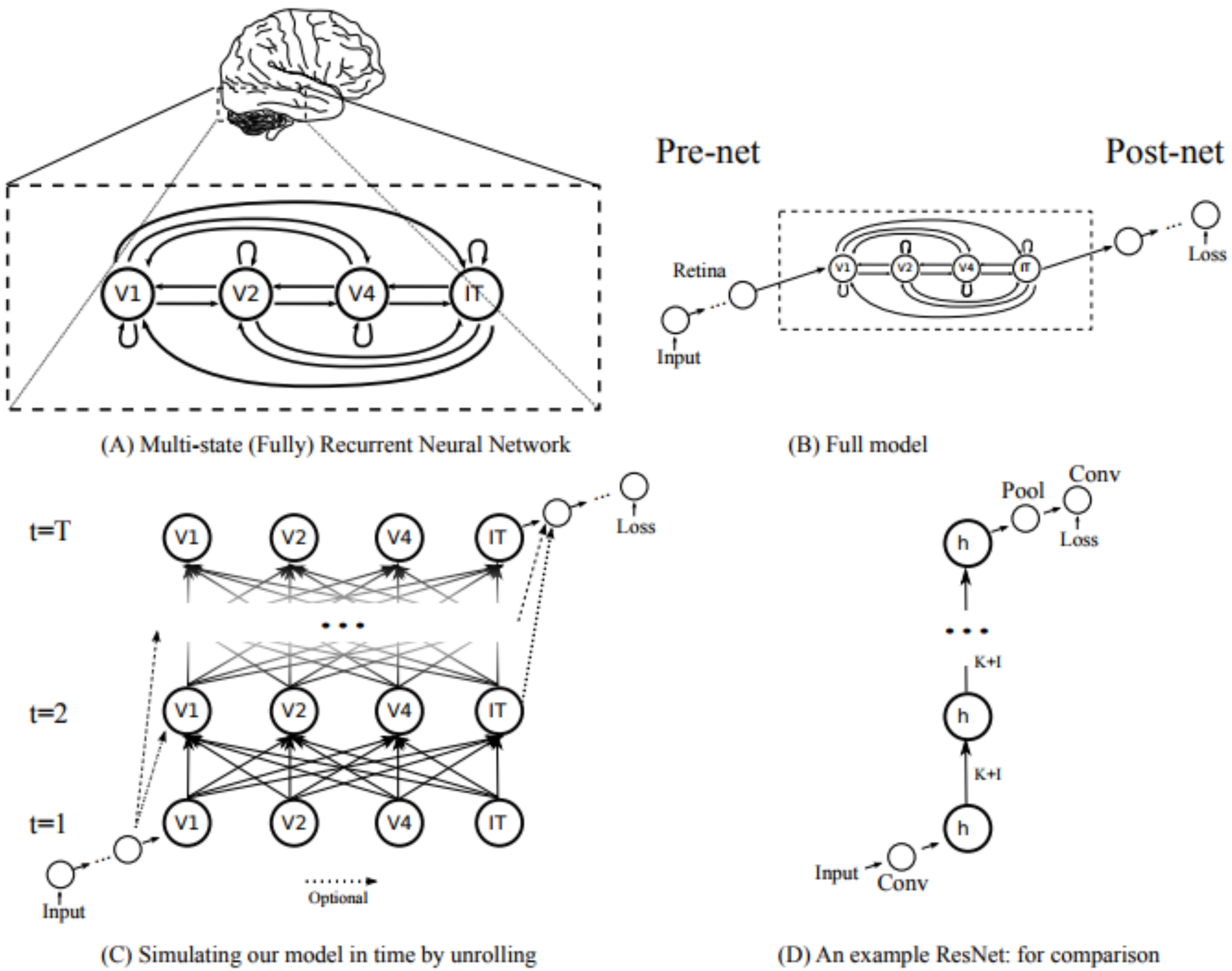


Figure 2: Modeling the ventral stream of visual cortex using a multi-state fully recurrent neural network

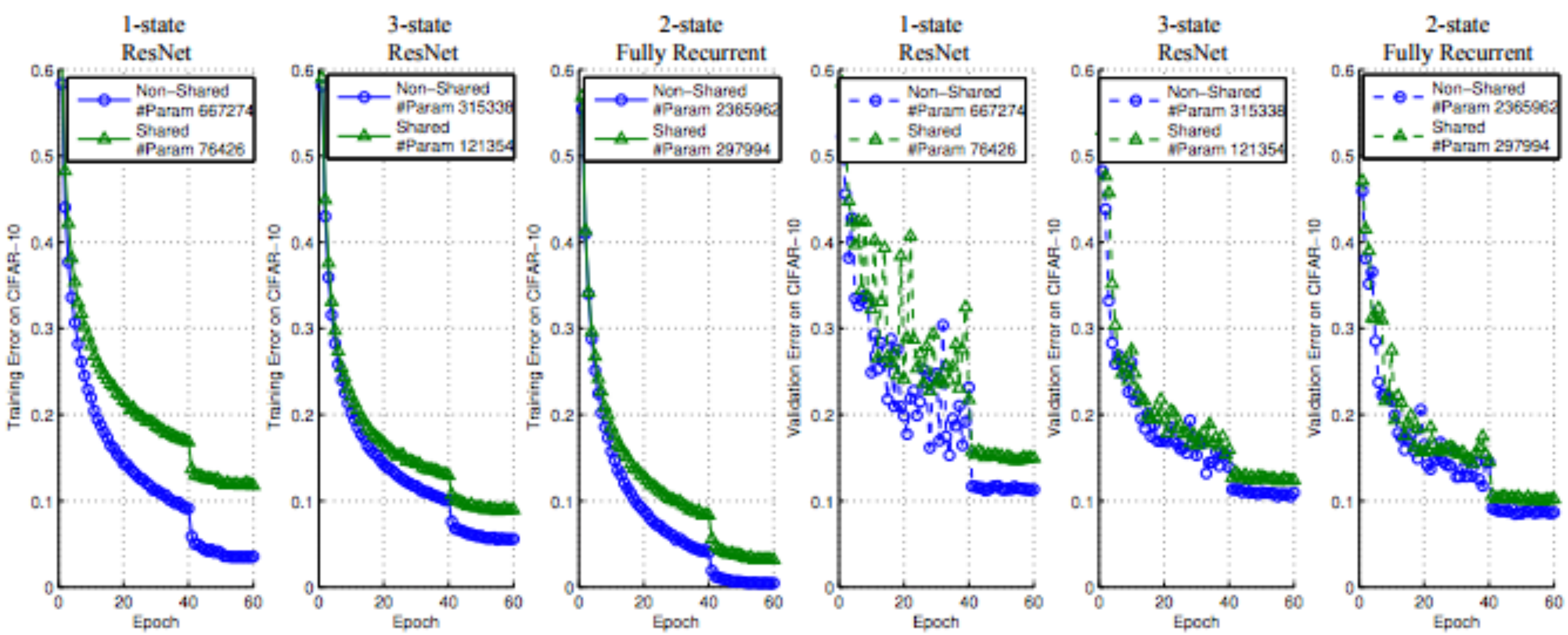
Models

Bridging the Gaps Between Residual Learning, Recurrent Neural Networks and Visual Cortex

by

Qianli Liao and Tomaso Poggio
Center for Brains, Minds and Machines, McGovern Institute, MIT

CIFAR-10 Error



Models

Feedback Networks

Amir R. Zamir[1,3*] Te-Lin Wu[1*] Lin Sun[1,2] William B. Shen[1] Bertram E. Shi[2]
Jitendra Malik[3] Silvio Savarese[1]
[1] Stanford University [2] HKUST [3] University of California, Berkeley
http://feedbacknet.stanford.edu/

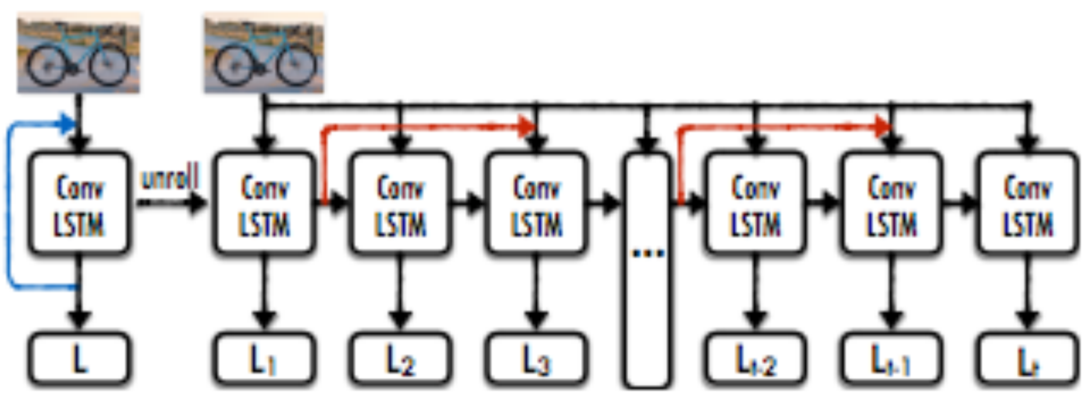


Figure 2. **Illustration of our core feedback model and skip connections (shown in red) when unrolled in time.** 'ConvLSTM' and 'L' boxes represent convolutional operations and iteration losses, respectively.

$$\mathbf{L} = \sum_{t=1}^{T} \gamma^{t} \mathbf{L_t}, \text{ where } \mathbf{L_t} = -log \frac{e^{\mathbf{H_t^D}[C]}}{\sum_j e^{\mathbf{H_t^D}[j]}}.$$

Feedback Networks

Amir R. Zamir[1,3*] Te-Lin Wu[1*] Lin Sun[1,2] William B. Shen[1] Bertram E. Shi[2]
Jitendra Malik[3] Silvio Savarese[1]

[1] Stanford University [2] HKUST [3] University of California, Berkeley
http://feedbacknet.stanford.edu/

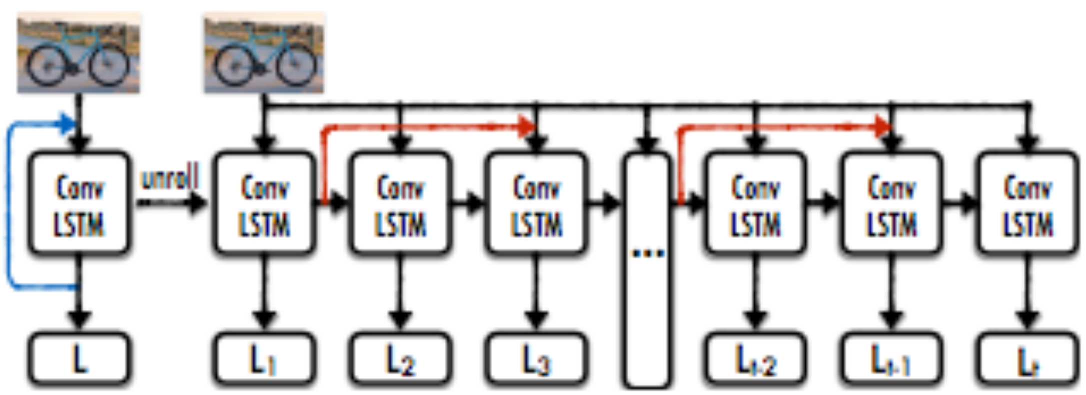


Figure 2. **Illustration of our core feedback model and skip connections (shown in red) when unrolled in time.** 'ConvLSTM' and 'L' boxes represent convolutional operations and iteration losses, respectively.

$$\mathbf{L} = \sum_{t=1}^{T} \gamma^t \mathbf{L_t}, \text{ where } \mathbf{L_t} = -log \frac{e^{\mathbf{H_t^D}[C]}}{\sum_j e^{\mathbf{H_t^D}[j]}}.$$

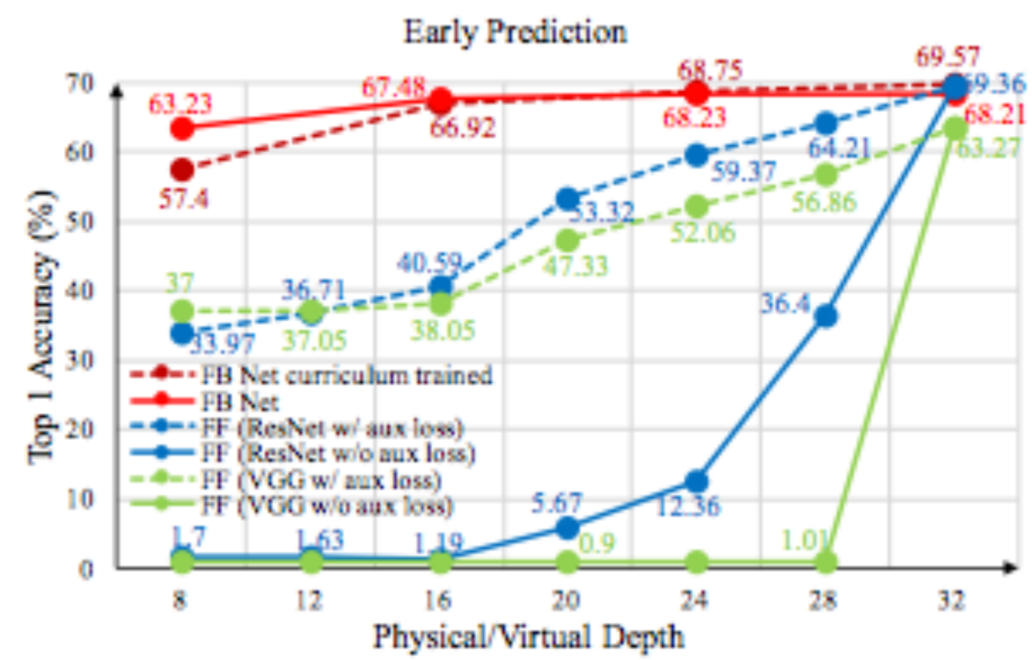


Figure 5. **Evaluation of early predictions.** Comparison of accuracy of feedback (FB) model and feedforward (FF) baselines (ResNet & VGG, with or without auxiliary loss layers)

Models

Feedback Networks

Amir R. Zamir[1,3*] Te-Lin Wu[1*] Lin Sun[1,2] William B. Shen[1] Bertram E. Shi[2]
Jitendra Malik[3] Silvio Savarese[1]

[1] Stanford University [2] HKUST [3] University of California, Berkeley
http://feedbacknet.stanford.edu/

Model	Physical Depth	Virtual Depth	Top1 (%)	Top5 (%)
Feedback Net	12	48	**71.12**	**91.51**
	8	32	69.57	91.01
	4	16	67.83	90.12
Feedforward (ResNet[19])	48	-	70.04	90.96
	32	-	69.36	91.07
	12	-	66.35	90.02
	8	-	64.23	88.95
	128*	-	70.92	91.28
	110*	-	72.06	92.12
	64*	-	71.01	91.48
	48*	-	70.56	91.60
	32*	-	69.58	91.55
Feedforward (VGG[48])	48	-	55.08	82.1
	32	-	63.56	88.41
	12	-	64.65	89.26
	8	-	63.91	88.90
Highway [53]	19	-	67.76	-
ResNet v2[20]	1001	-	77.29	-
Stochastic Depth [24]	110	-	75.02	-
SwapOut [49]	32 fat	-	77.28	-
RCNN [37]	4 fat	16	68.25	-

Table 6. **Endpoint performance comparison on CIFAR-100.** Baselines denoted with * are the architecture used in the original ResNet paper.

$$\mathbf{L} = \sum_{t=1}^{T} \gamma^t \mathbf{L_t}, \text{ where } \mathbf{L_t} = -log \frac{e^{\mathbf{H}_t^{\mathbf{D}}[C]}}{\sum_j e^{\mathbf{H}_t^{\mathbf{D}}[j]}}.$$

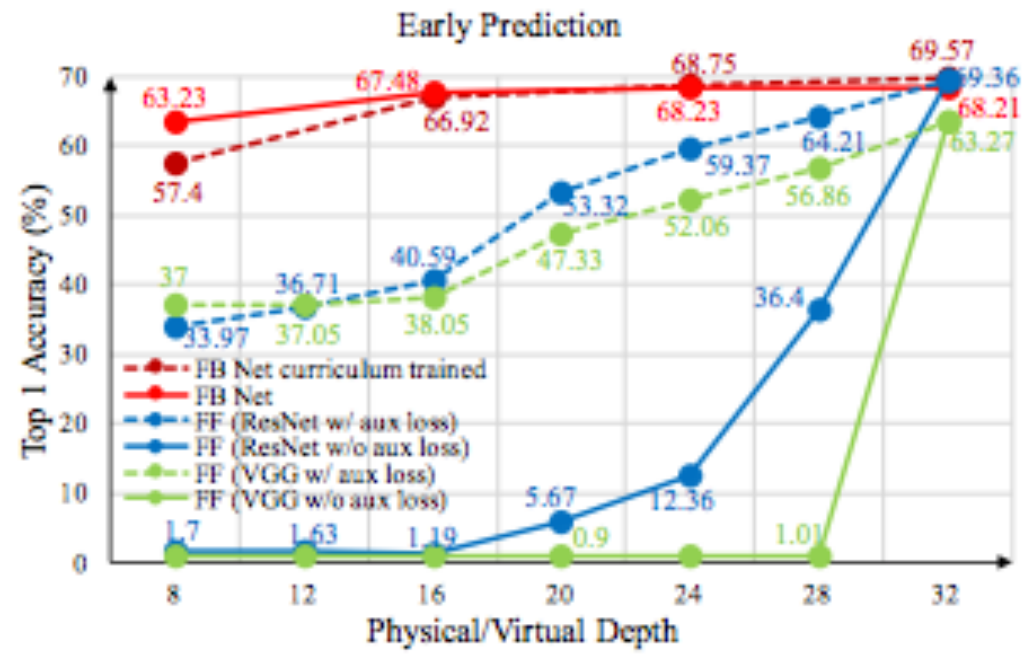


Figure 5. **Evaluation of early predictions.** Comparison of accuracy of feedback (FB) model and feedforward (FF) baselines (ResNet & VGG, with or without auxiliary loss layers)

Models

Feedback Networks

Amir R. Zamir[1,3*] Te-Lin Wu[1*] Lin Sun[1,2] William B. Shen[1] Bertram E. Shi[2]
Jitendra Malik[3] Silvio Savarese[1]

[1] Stanford University [2] HKUST [3] University of California, Berkeley
http://feedbacknet.stanford.edu/

Model	Physical Depth	Virtual Depth	Top1 (%)	Top5 (%)
Feedback Net	12	48	**71.12**	**91.51**
	8	32	69.57	91.01
	4	16	67.83	90.12
Feedforward (ResNet[19])	48	-	70.04	90.96
	32	-	69.36	91.07
	12	-	66.35	90.02
	8	-	64.23	88.95
	128*	-	70.92	91.28
	110*	-	72.06	92.12
	64*	-	71.01	91.48
	48*	-	70.56	91.60
	32*	-	69.58	91.55
Feedforward (VGG[48])	48	-	55.08	82.1
	32	-	63.56	88.41
	12	-	64.65	89.26
	8	-	63.91	88.90
Highway [53]	19	-	67.76	-
ResNet v2[20]	1001	-	77.29	-
Stochastic Depth [24]	110	-	75.02	-
SwapOut [49]	32 fat	-	77.28	-
RCNN [37]	4 fat	16	68.25	-

Table 6. **Endpoint performance comparison on CIFAR-100.** Baselines denoted with * are the architecture used in the original ResNet paper.

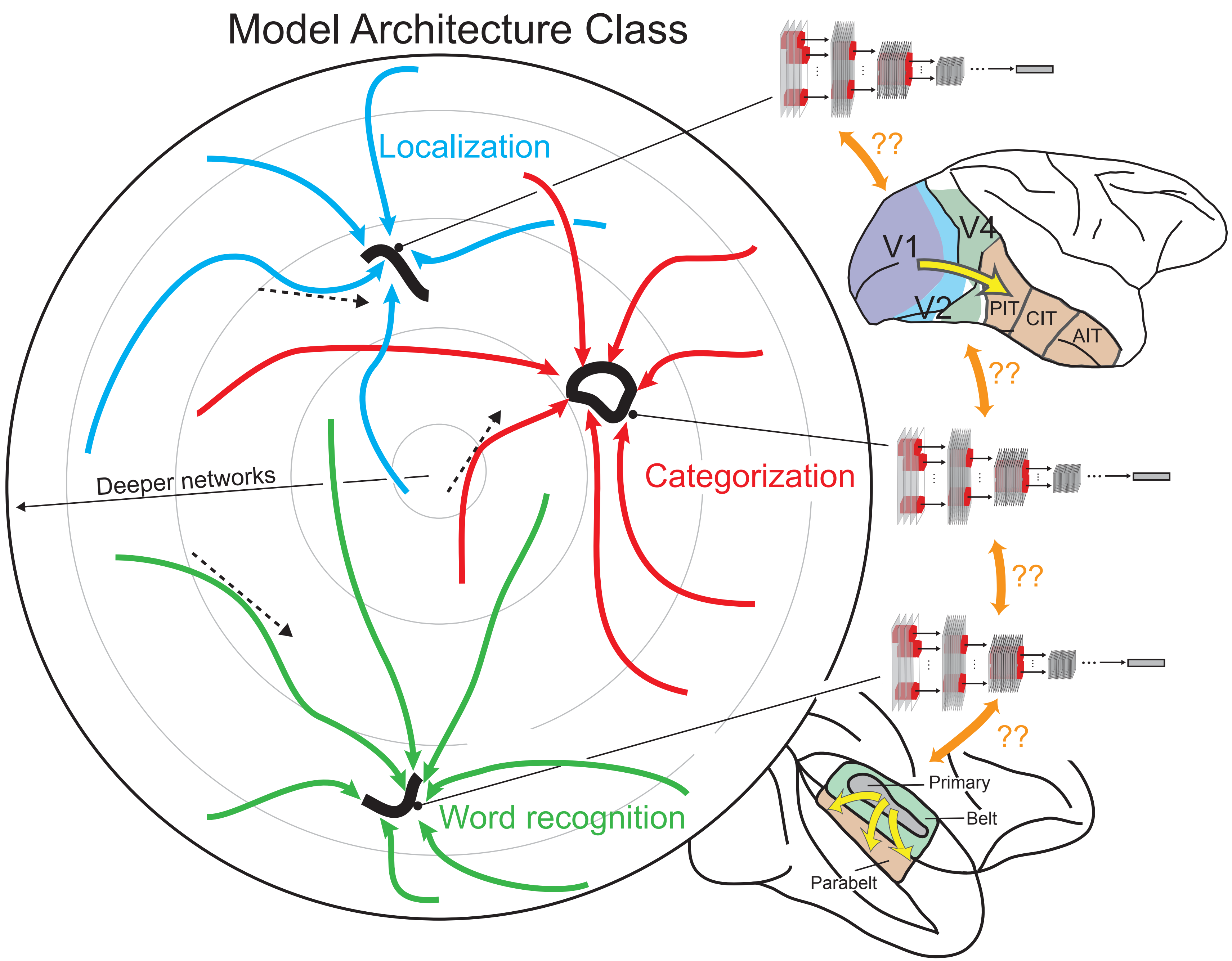
$$\mathbf{L} = \sum_{t=1}^{T} \gamma^t \mathbf{L}_t, \text{ where } \mathbf{L}_t = -log \frac{e^{\mathbf{H}_t^{\mathrm{D}}[C]}}{\sum_j e^{\mathbf{H}_t^{\mathrm{D}}[j]}}.$$

> Formulate comprehensive model class (**CNNs**)

> Choose challenging, ethologically-valid tasks (**categorization**)

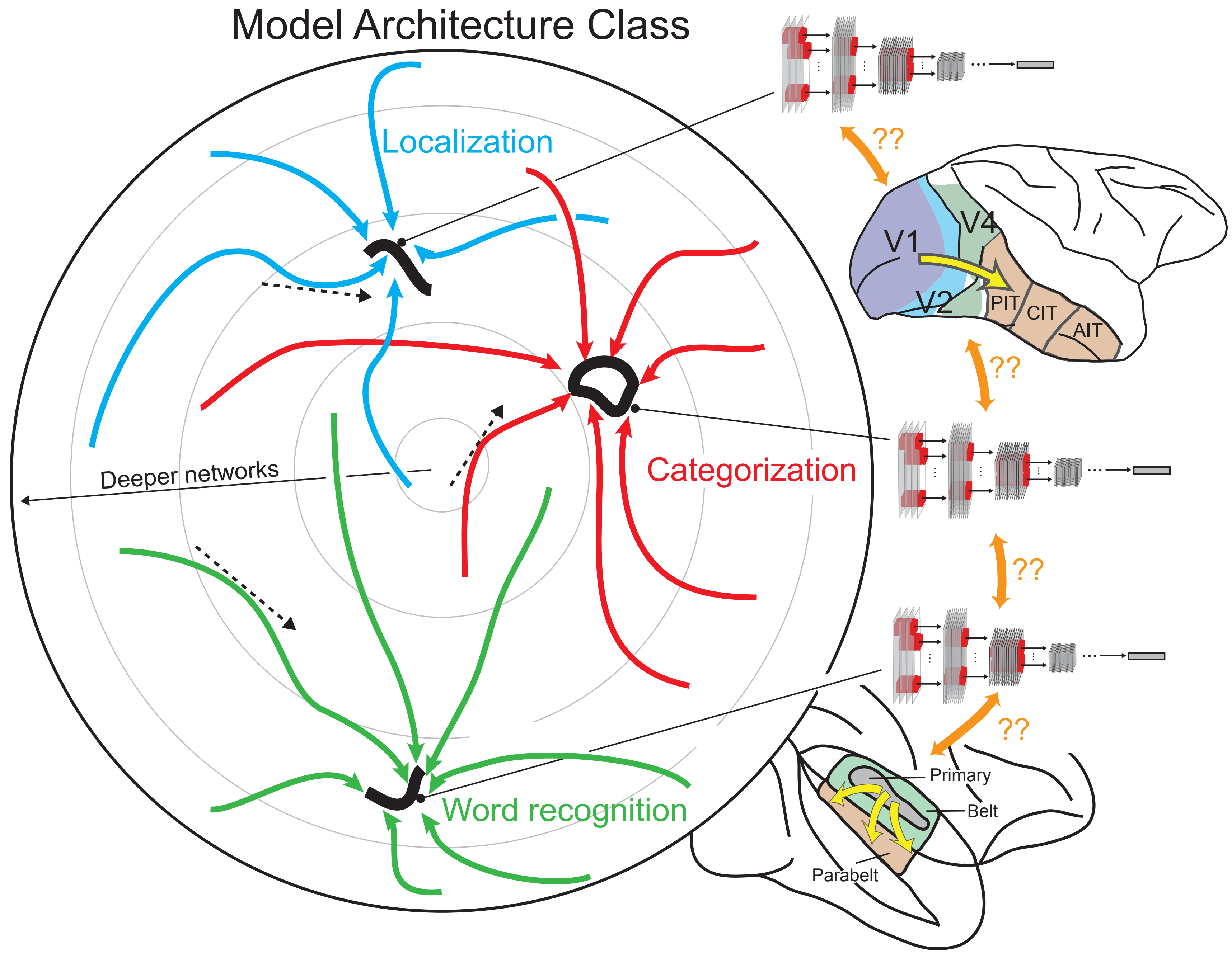
> Implement generic learning rules (**gradient descent**)

> Map to brain data. (**temporal averages in ventral stream**)



> Formulate comprehensive model class (**ConvRNNs**)

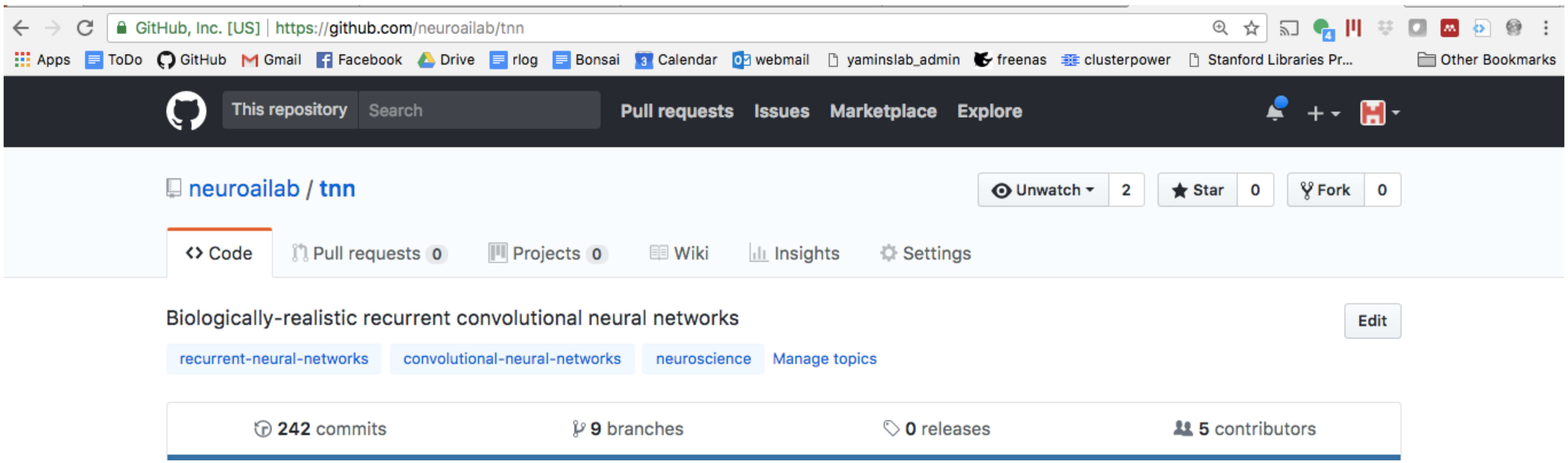
> Choose challenging, ethologically-valid tasks (**??**)

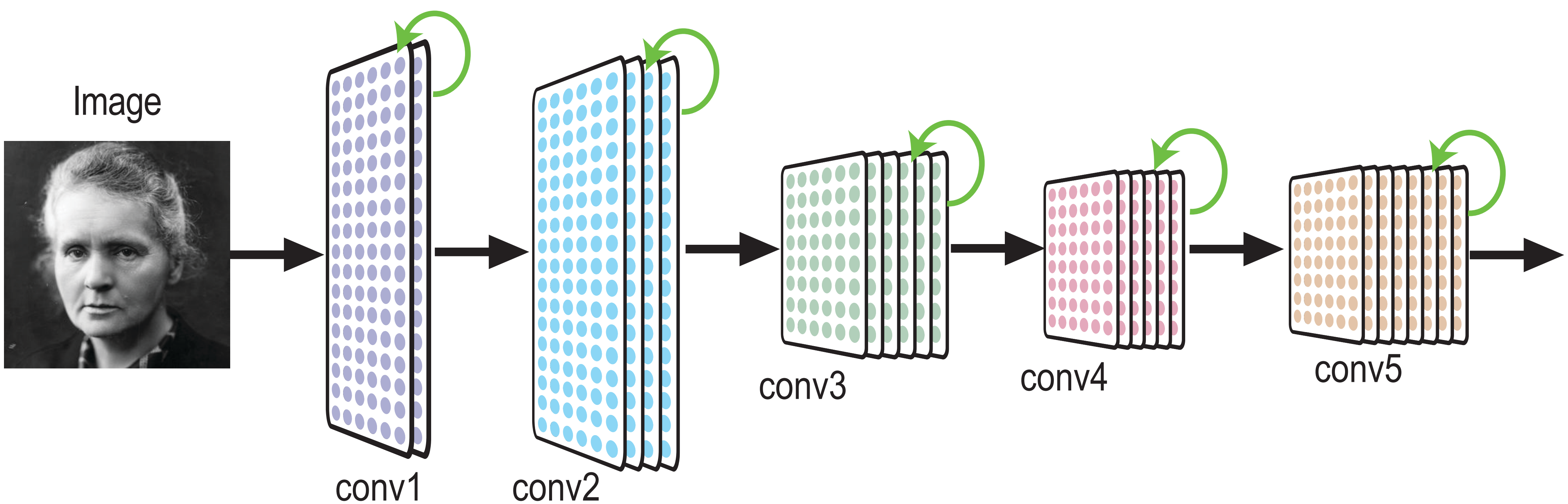


> Map to brain data. (**ventral stream dynamics**)

Temporal Neural Networks (TNN) Library

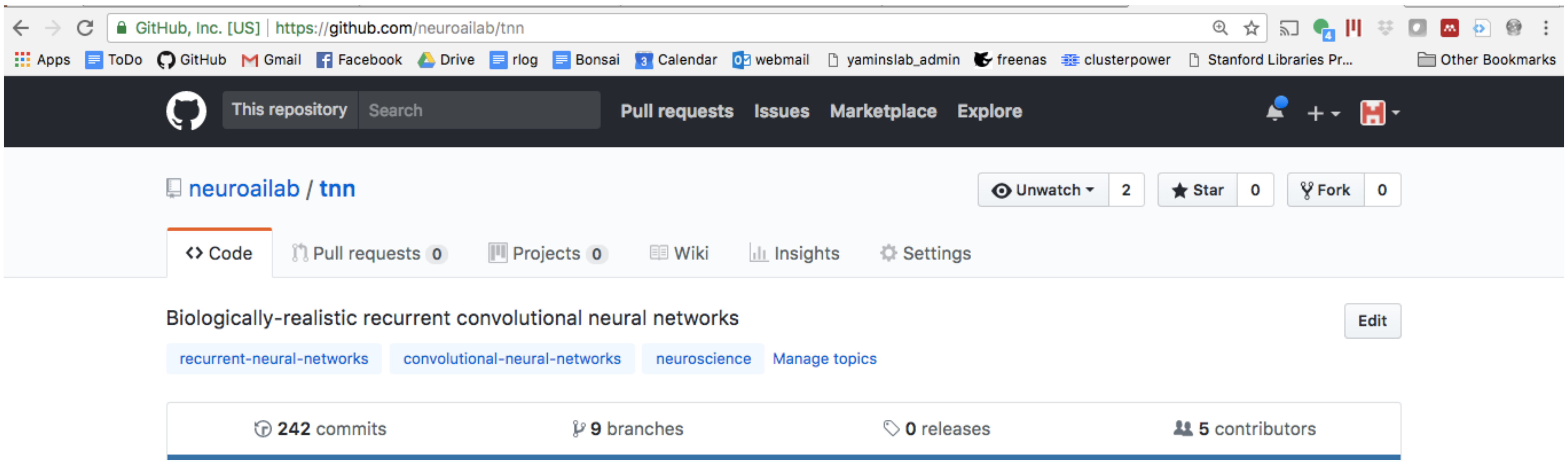
http://github.com/neuroailab/tnn

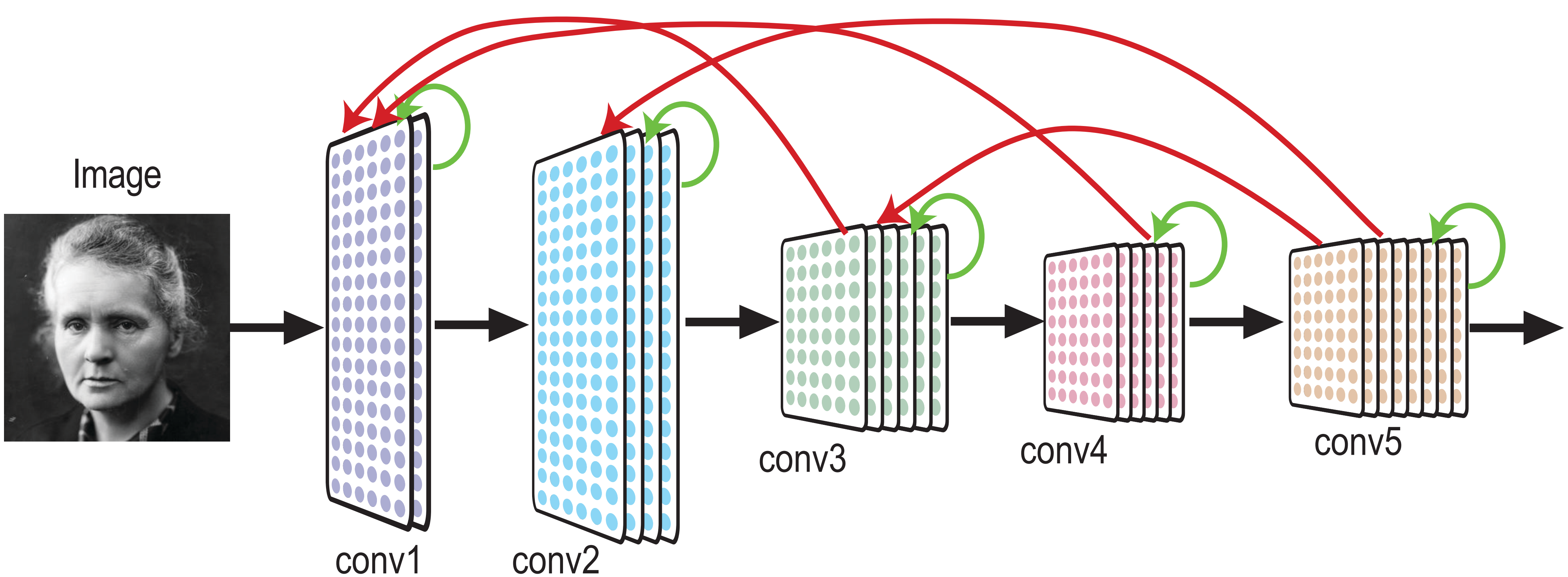




Temporal Neural Networks (TNN) Library

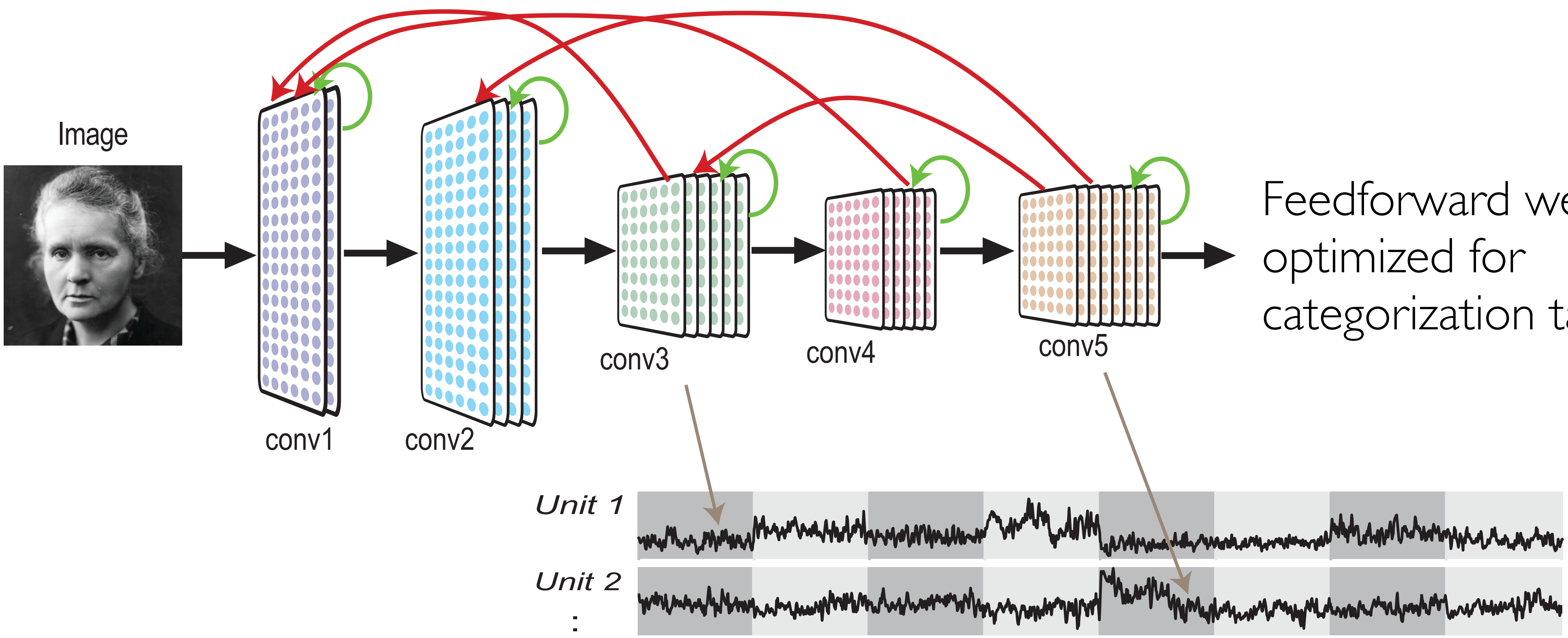
http://github.com/neuroailab/tnn





Fitting Recurrent Dynamics Directly

Convolutional RNNs (convRNNs) with **local** and **long-range** feedback:

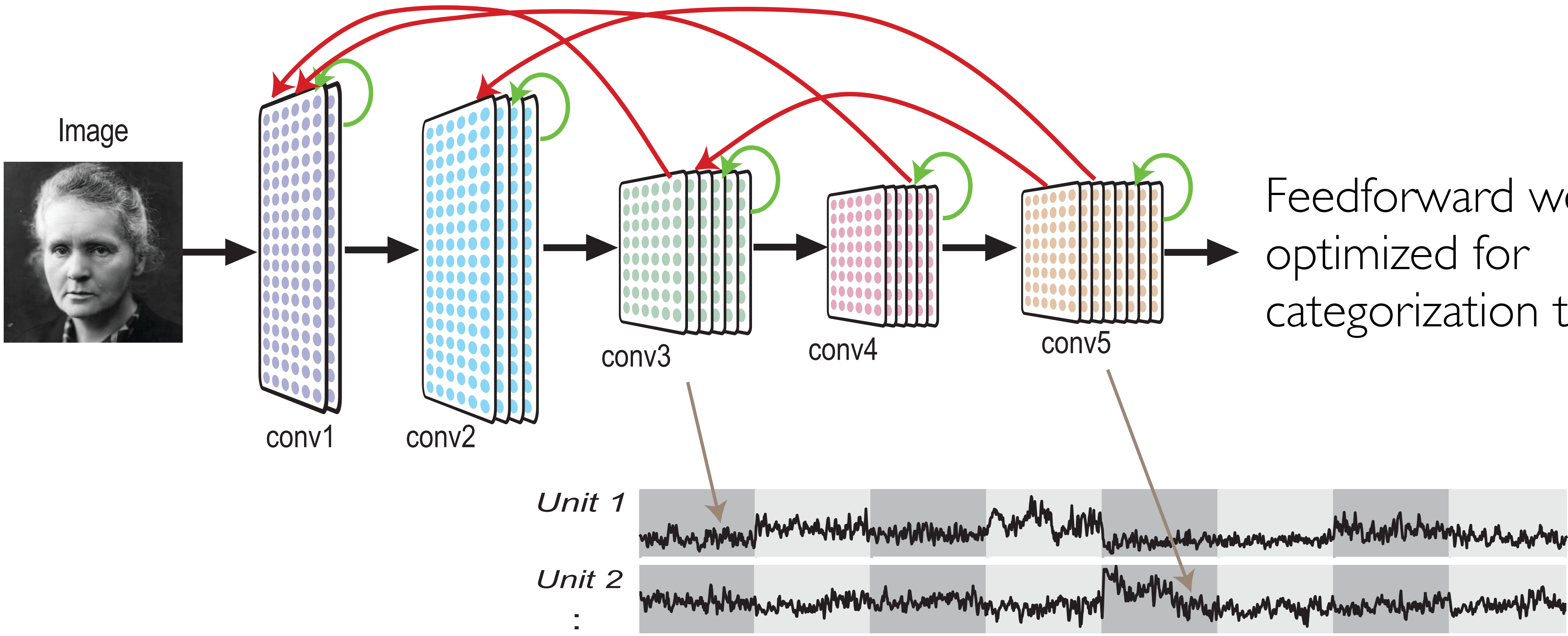


Feedforward weights optimized for categorization task

Recurrent weights optimized to match neural dynamics in V4 and IT

Fitting Recurrent Dynamics Directly

Convolutional RNNs (convRNNs) with **local** and **long-range** feedback:

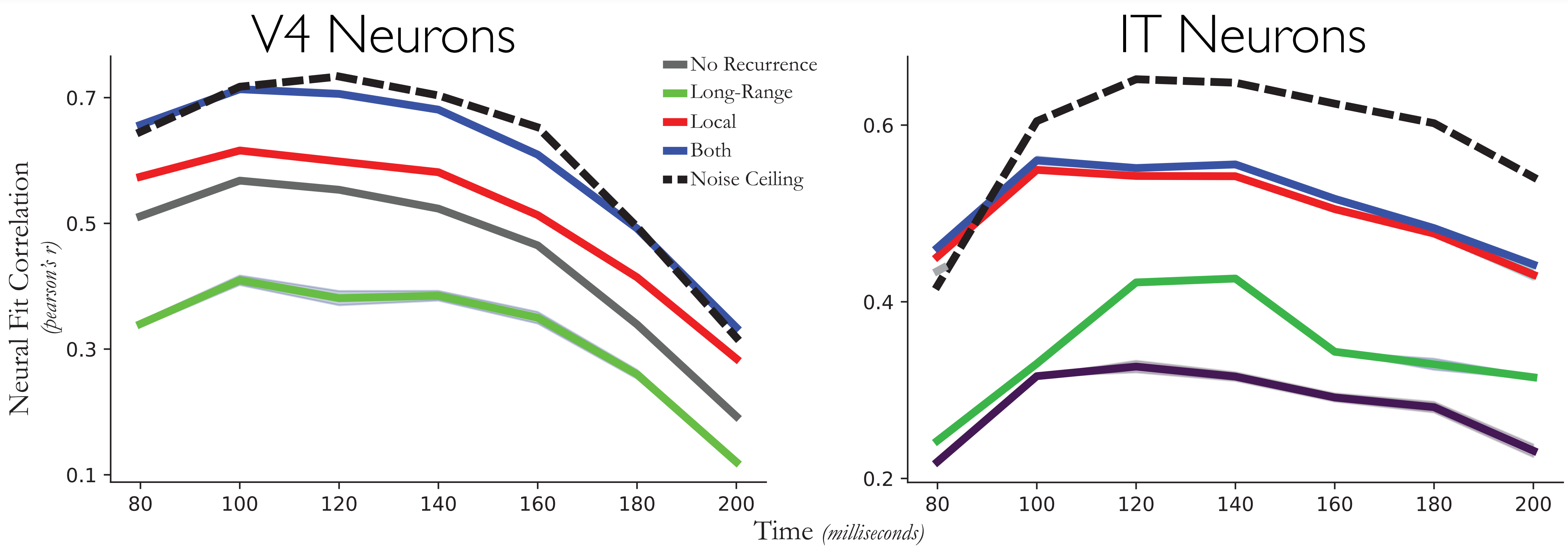


Feedforward weights optimized for categorization task

Recurrent weights optimized to match neural dynamics in V4 and IT

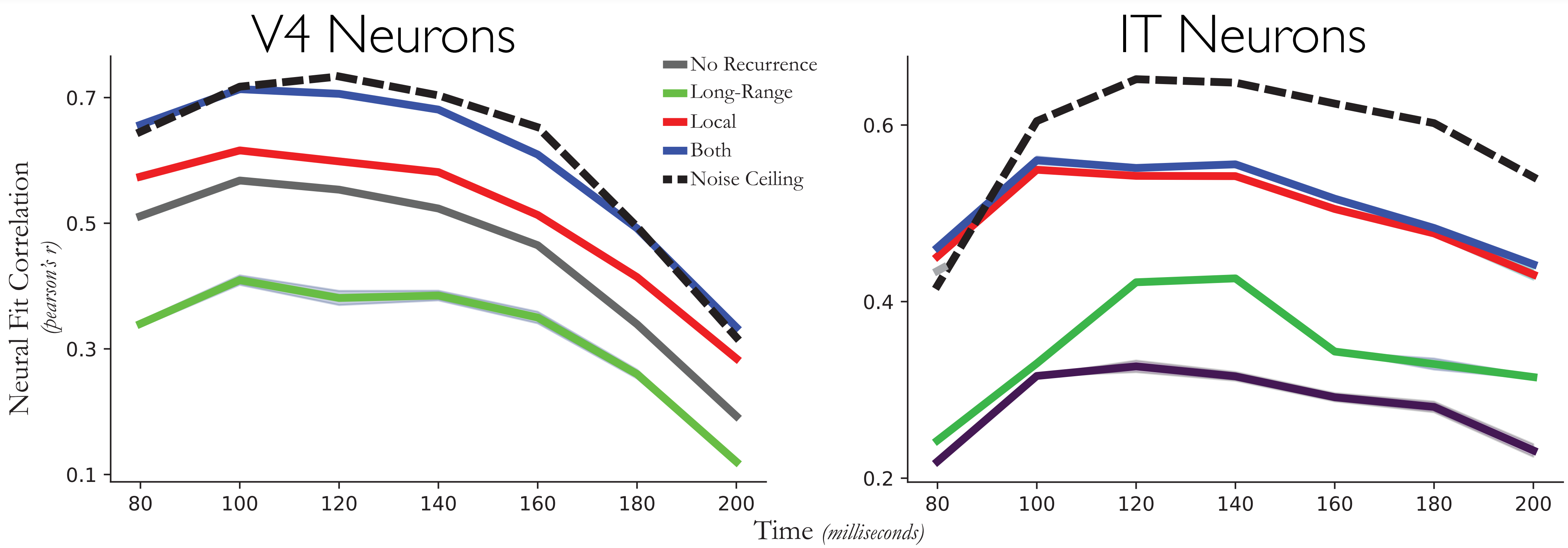
Loss: L2 averaged over 10ms timebins up to 250ms

Fitting Recurrent Dynamics Directly



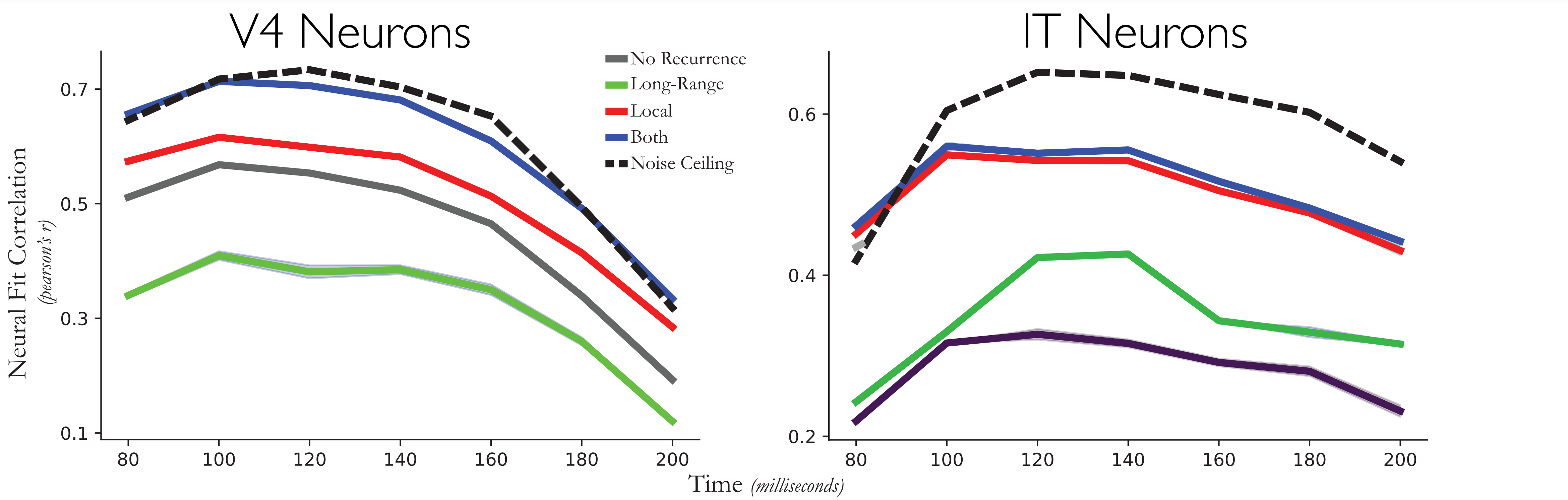
- Local recurrent circuits substantially improves predictions of IT dynamics

Fitting Recurrent Dynamics Directly

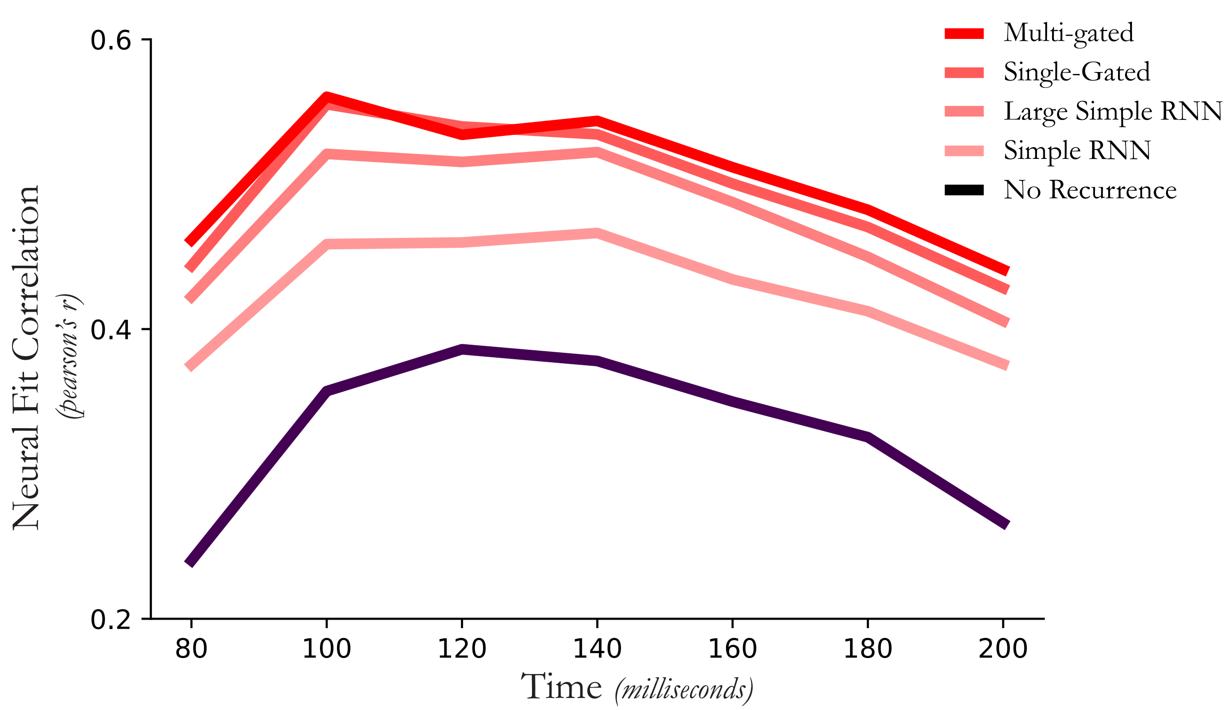


- Local recurrent circuits substantially improves predictions of IT dynamics
- Long-range feedback improves V4 predictions nearly to 100% of noise ceiling.

Fitting Recurrent Dynamics Directly

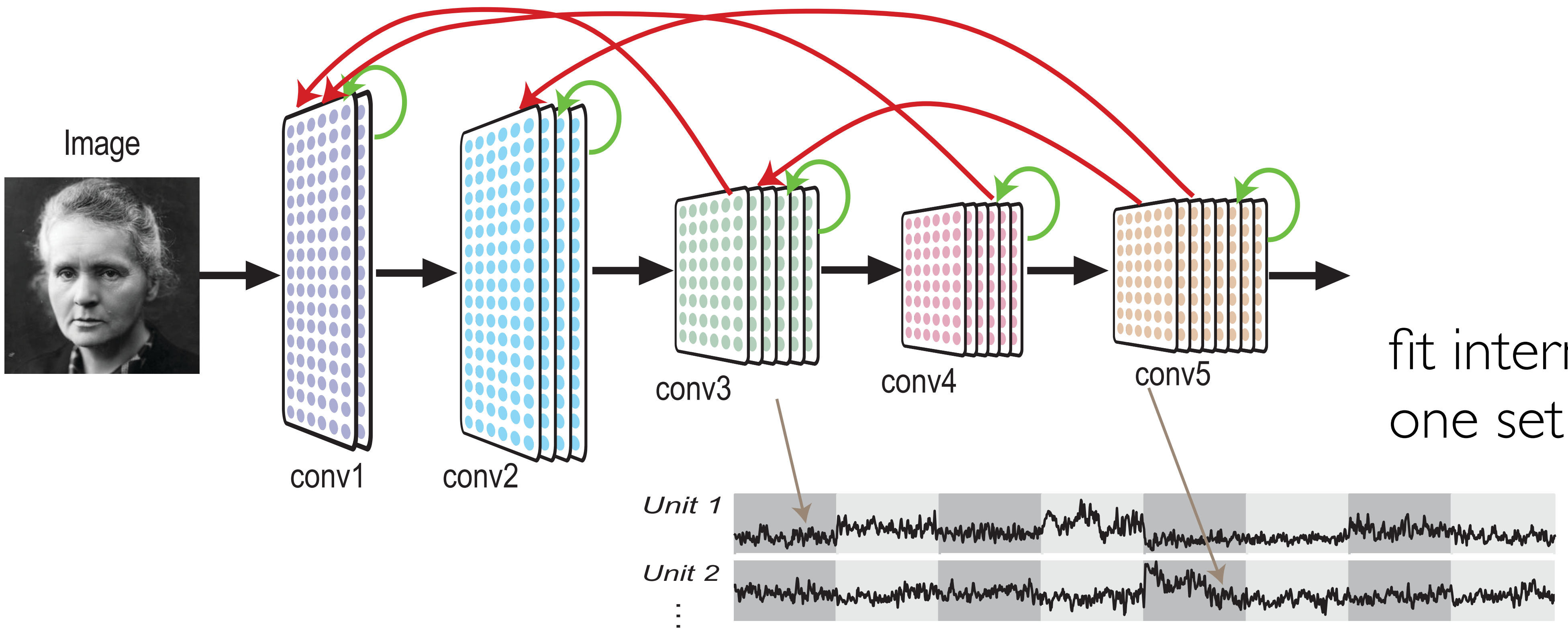


- Local recurrent circuits substantially improves predictions of IT dynamics
- Long-range feedback improves V4 predictions nearly to 100% of noise ceiling.
- Gating important for improved neural fit.



Fitting Recurrent Dynamics Directly

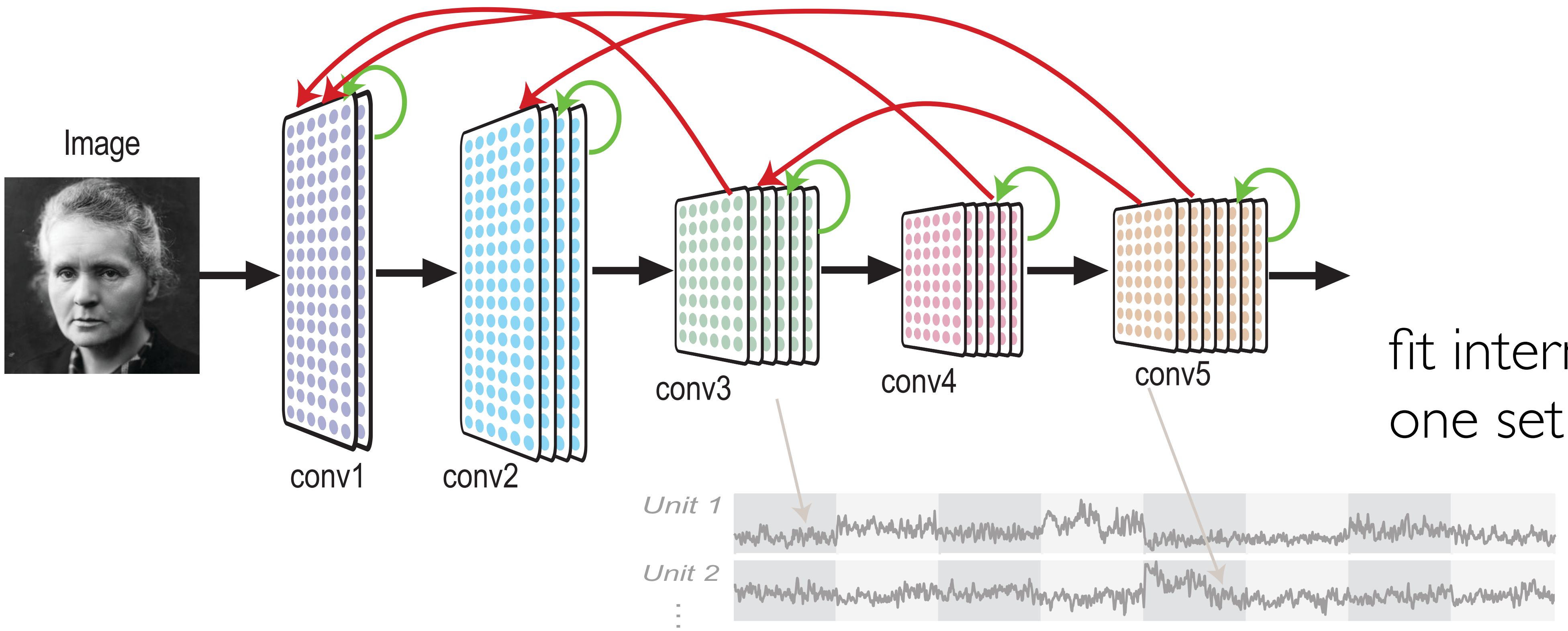
Fits hold pretty well even for held-out neuron cross-validation (as well as cross-image)



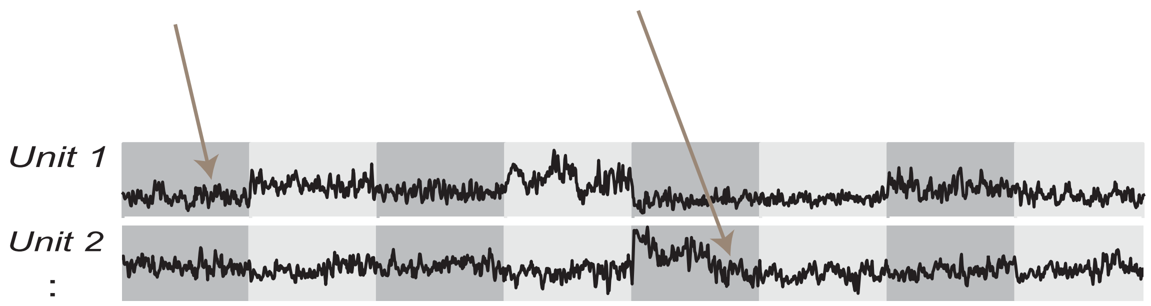
fit internal recurrences w/ one set of neurons

Fitting Recurrent Dynamics Directly

Fits hold pretty well even for held-out neuron cross-validation (as well as cross-image)



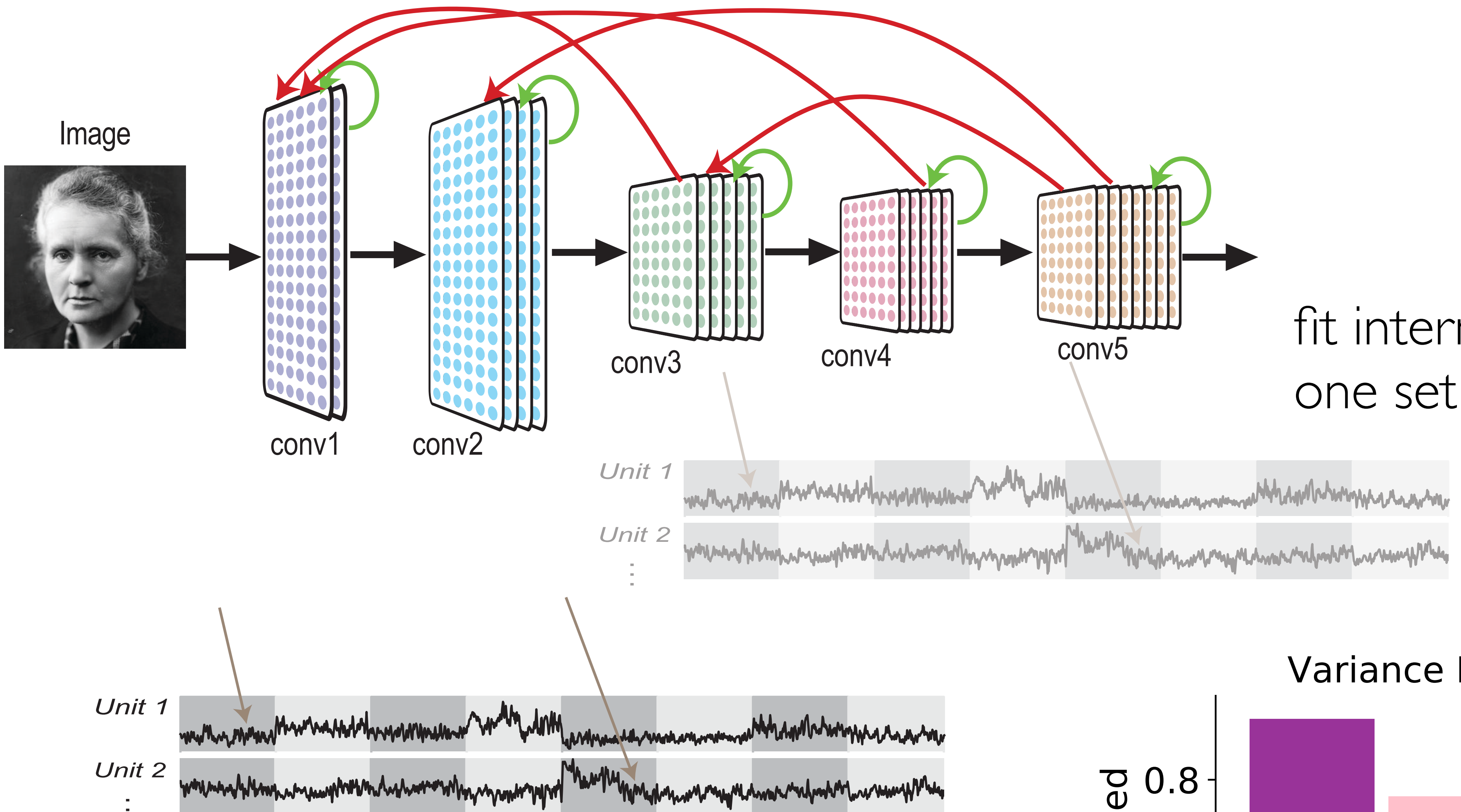
fit internal recurrences w/ one set of neurons



test on another with just linear regression fit

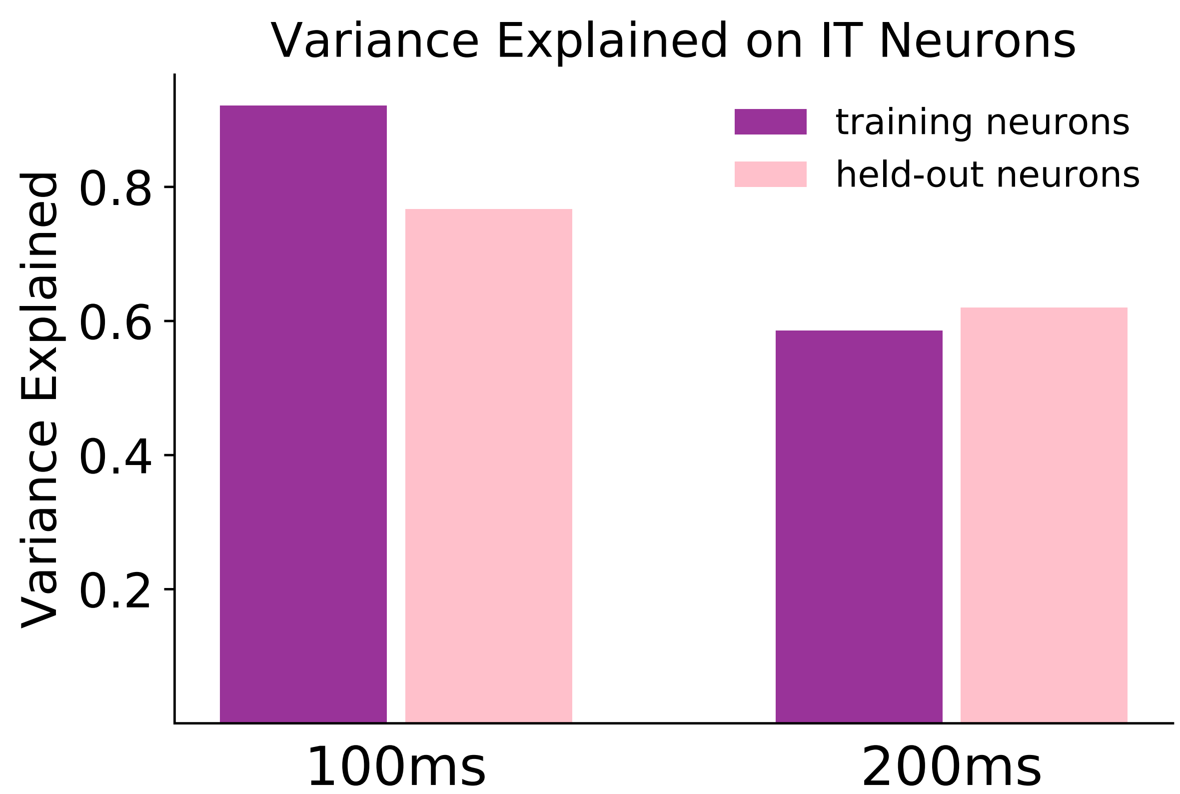
Fitting Recurrent Dynamics Directly

Fits hold pretty well even for held-out neuron cross-validation (as well as cross-image)



fit internal recurrences w/ one set of neurons

test on another with just linear regression fit



Fitting Recurrent Dynamics Directly

1.

A = *architecture class*

CNNs -> RNNs

"task"

2.

L = *loss function* ***D*** = *dataset*

e.g. **Object Categorization**

3. **Learning Rule**

$$\underset{a \in \mathcal{A}}{\text{argmin}}[L(p_a^*)]$$

architecture search

where p^* is result of

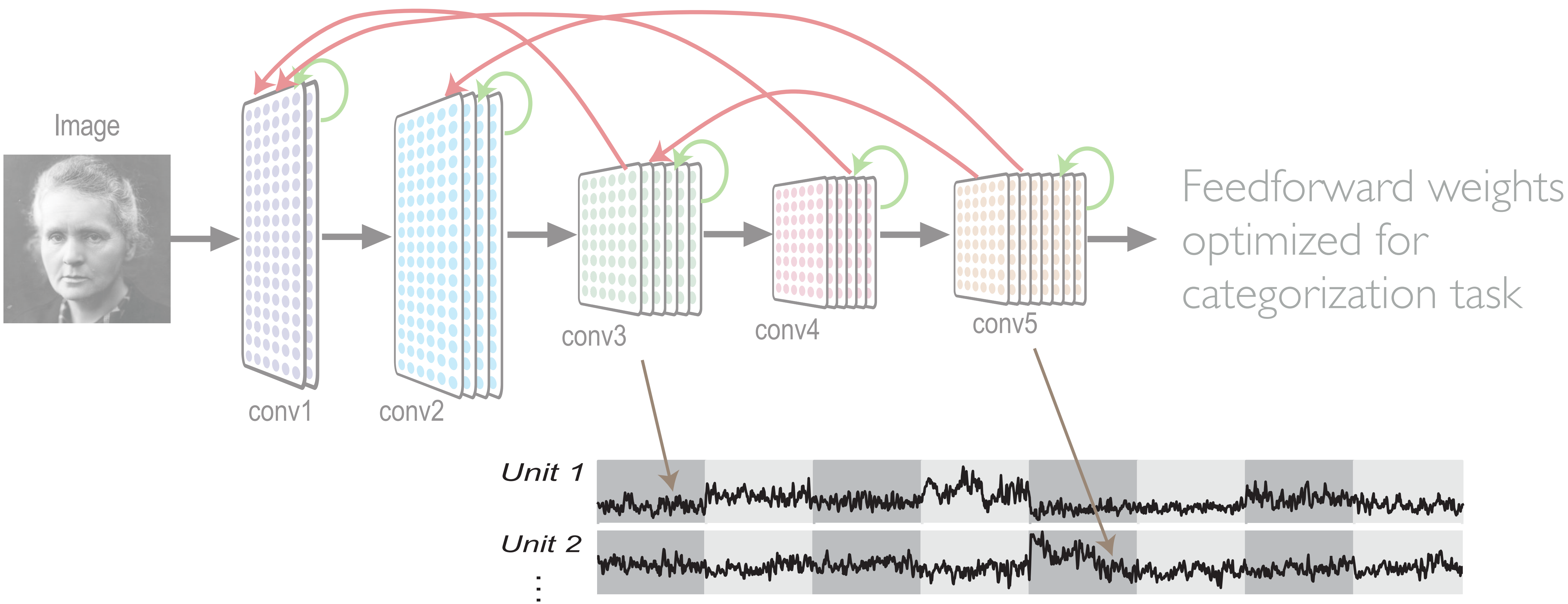
backprop

$$\frac{dp_a}{dt} = -\lambda(t) \cdot \langle \nabla_{p_a} L(x) \rangle_{x \in \mathcal{D}}$$

e.g. **Gradient Descent via Backprop**

Fitting Recurrent Dynamics Directly

Convolutional RNNs with **local** and **long-range** feedback:



Recurrent weights <u>optimized to match neural dynamics</u> in V4 and IT

Not a normative theory — no task.

Task-Driven Models?

1.

A *= architecture class*

CNNs -> RNNs

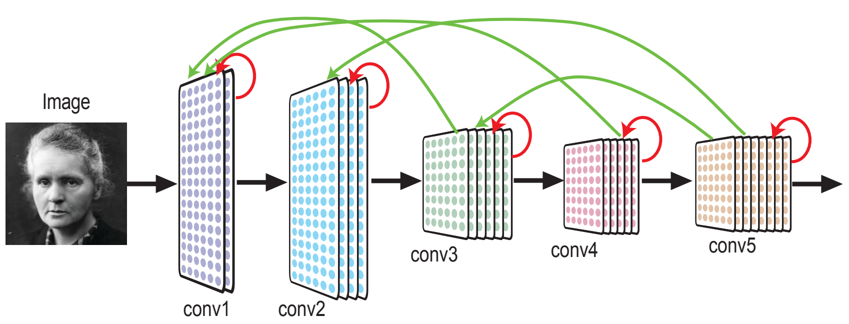
2.

"task"

L *= loss function* ***D*** *= dataset*

e.g. **Object Categorization**

What task(s) explain recurrences??



Task-Driven Models?

1.

A = architecture class

CNNs -> RNNs

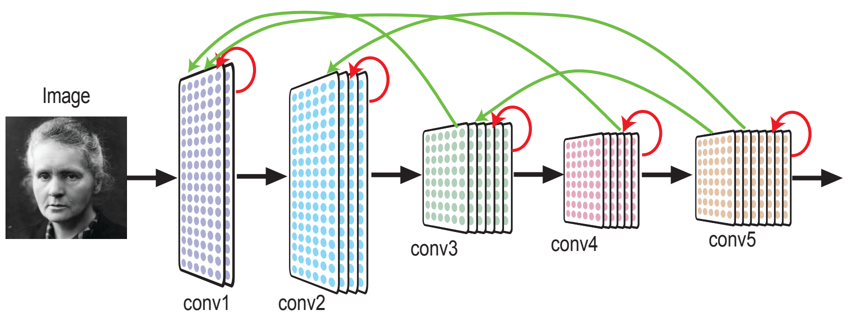
"task"

2.

L = loss function *D = dataset*

e.g. **Object Categorization**

What task(s) explain recurrences??



Hard Images
(e.g. heavy occlusion)

Time-accuracy tradeoff
(be correct but fast)

Temporal Goal
(e.g. motion-based)

Task-Driven Models?

“task”

1.

A = *architecture class*

CNNs -> RNNs

2.

L = *loss function* ***D*** = *dataset*

e.g. **Object Categorization**

3. **Learning Rule**

$$\underset{a \in \mathcal{A}}{\mathrm{argmin}}[L(p_a^*)]$$

architecture search

where p^* *is result of*

backprop

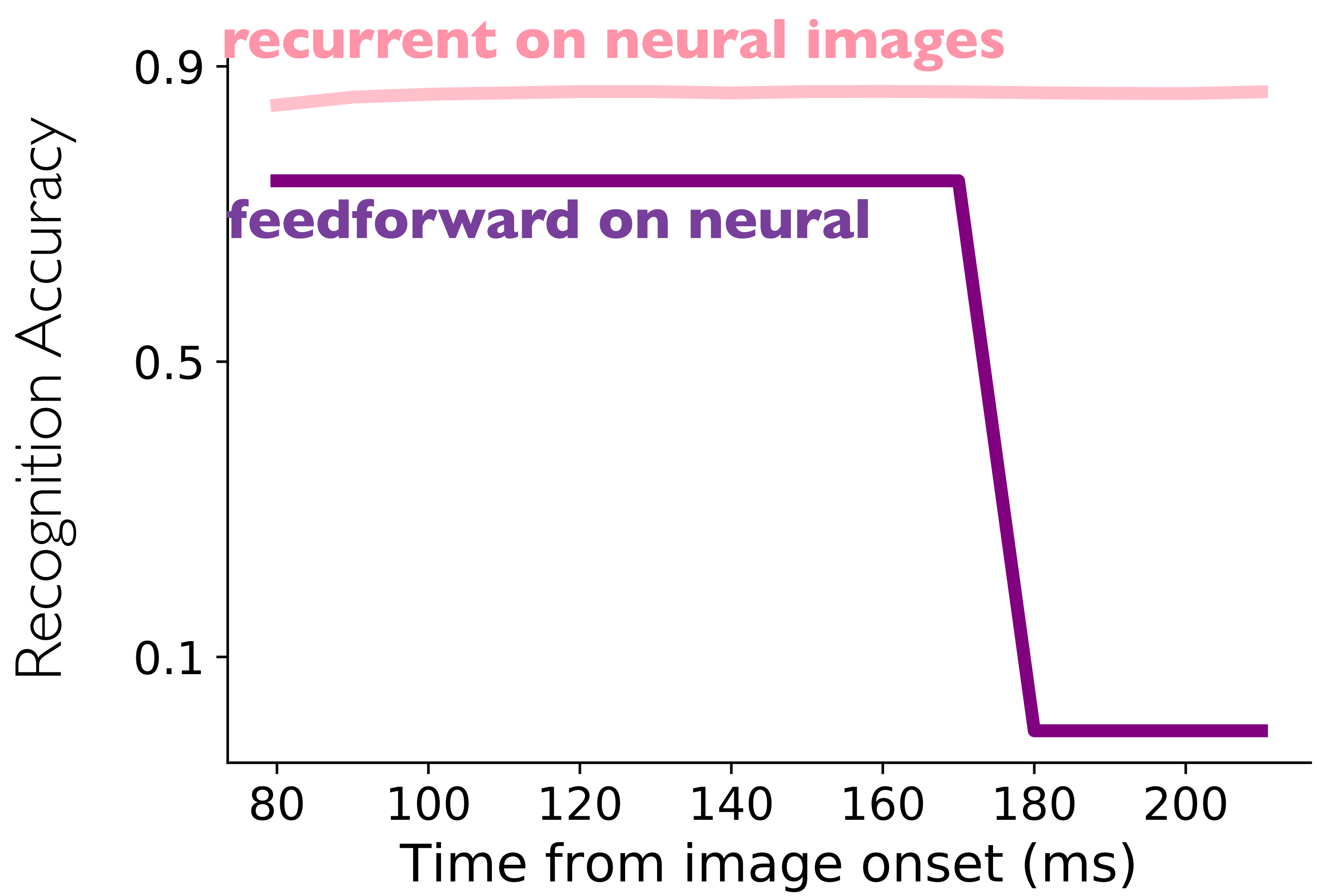
$$\frac{dp_a}{dt} = -\lambda(t) \cdot \langle \nabla_{p_a} L(x) \rangle_{x \in \mathcal{D}}$$

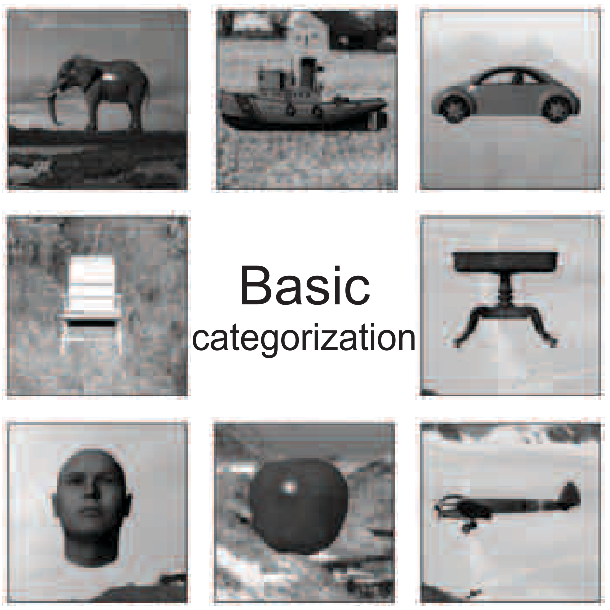
(possibility: actually, recurrence not used on-line)

(e.g. “just” implementing learning)

Task-Driven Models?

Performance on recognition task on neural images is improved 😀

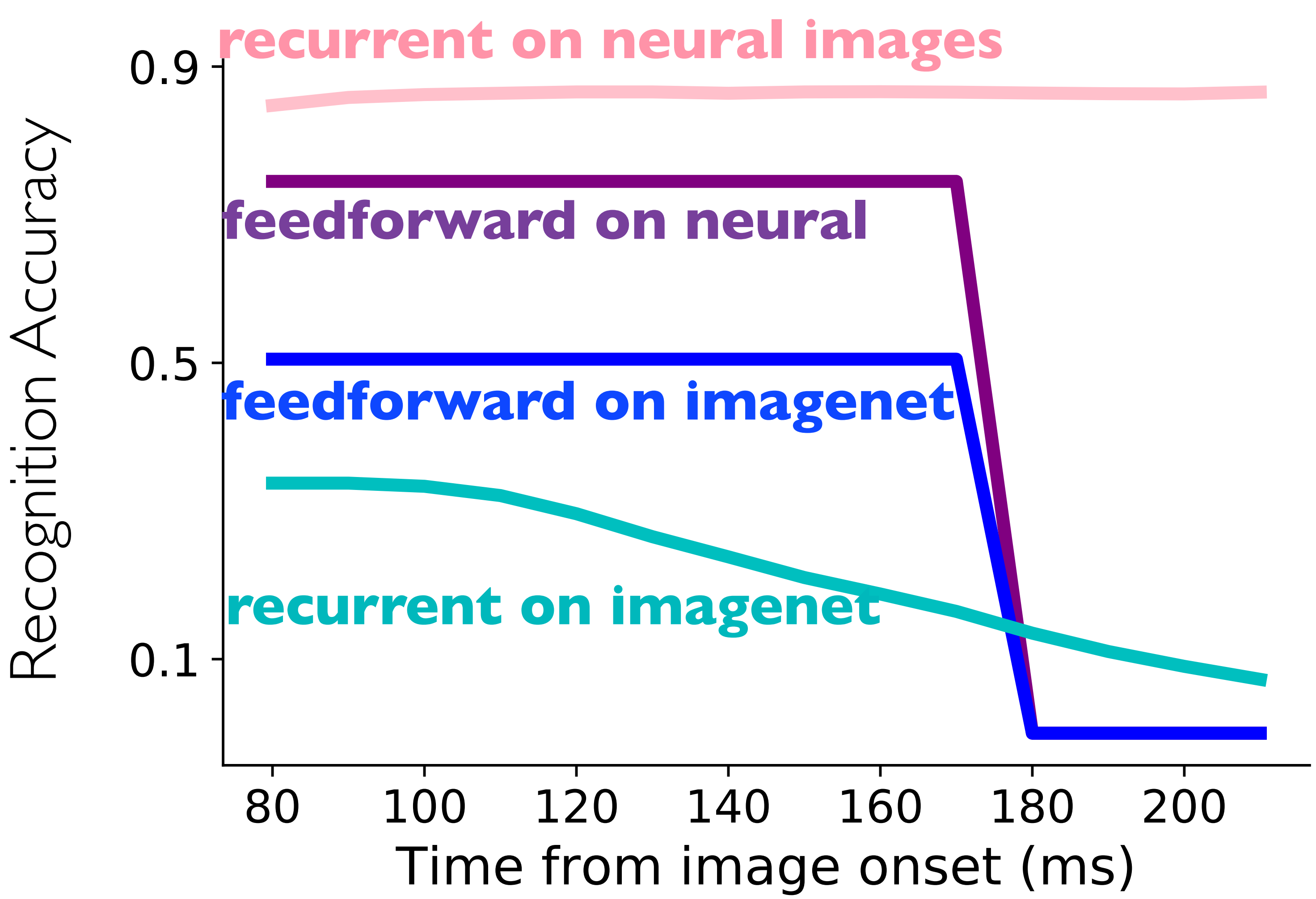


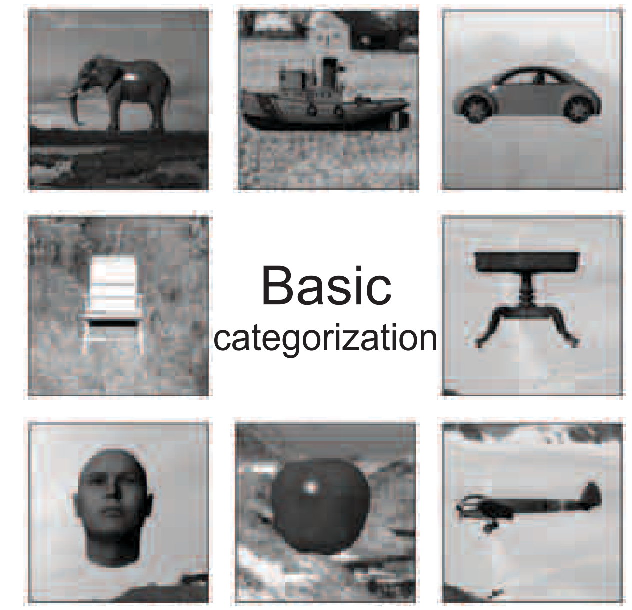


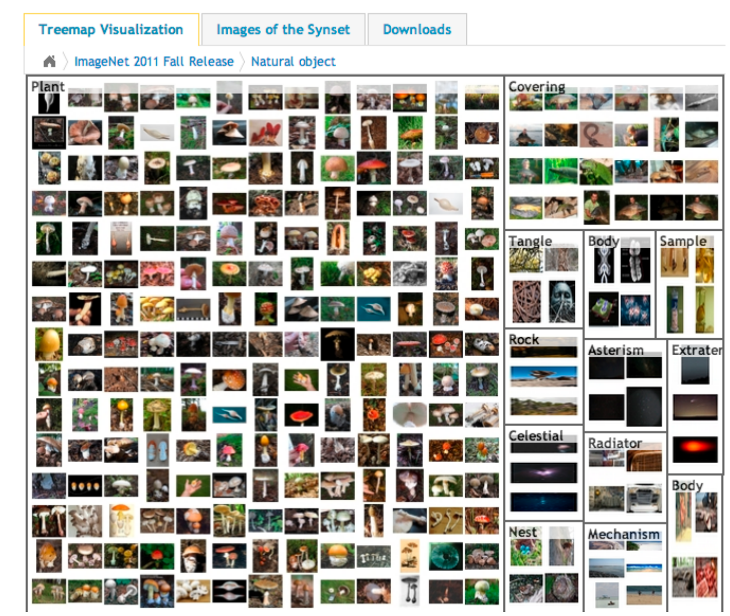
Task-Driven Models?

Performance on recognition task on neural images is improved 😀

. . . but performance on Imagenet dramatically worsens. 🙁 🙁

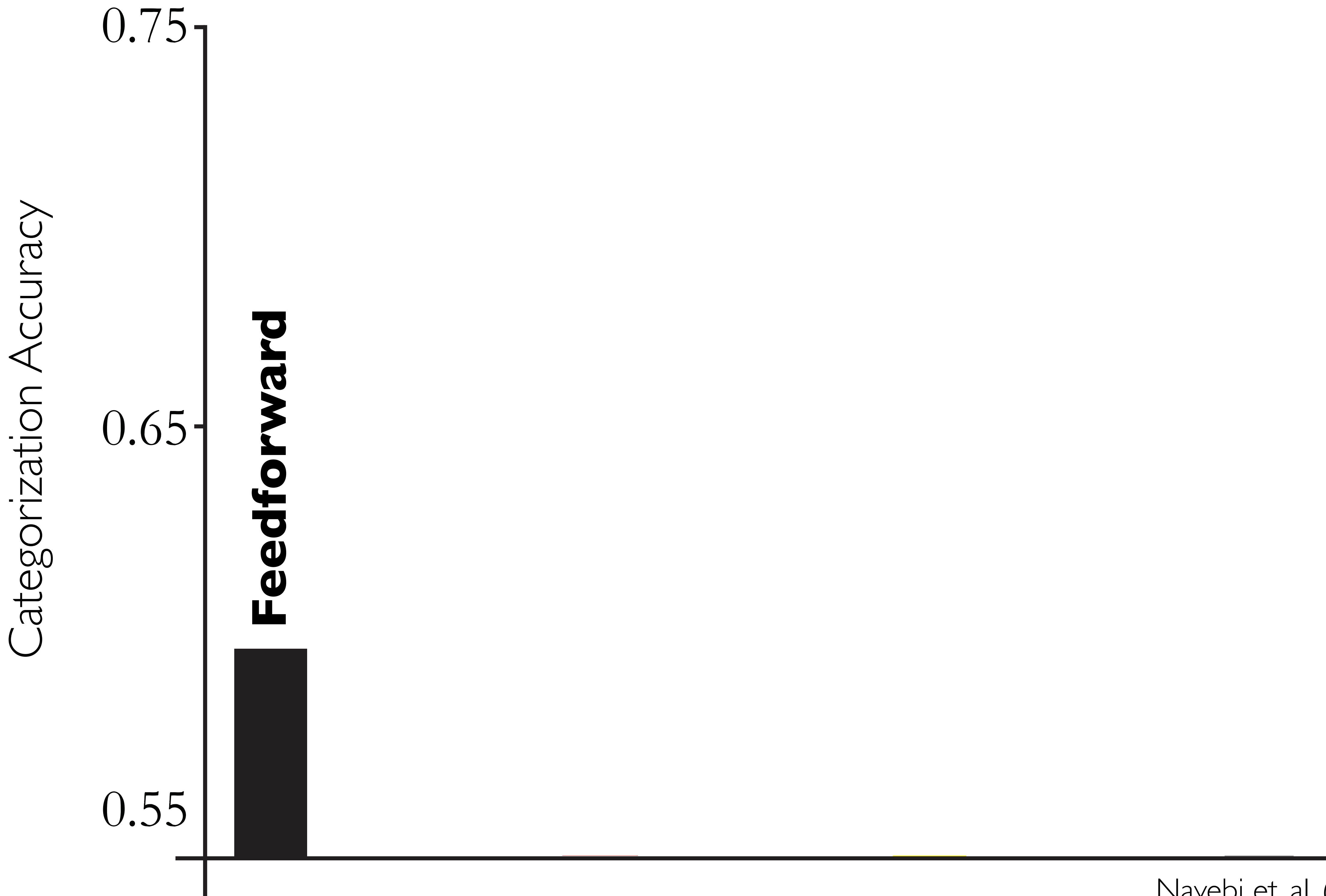






⟹ subtle "overfitting" to image-type or animal idiosyncracies

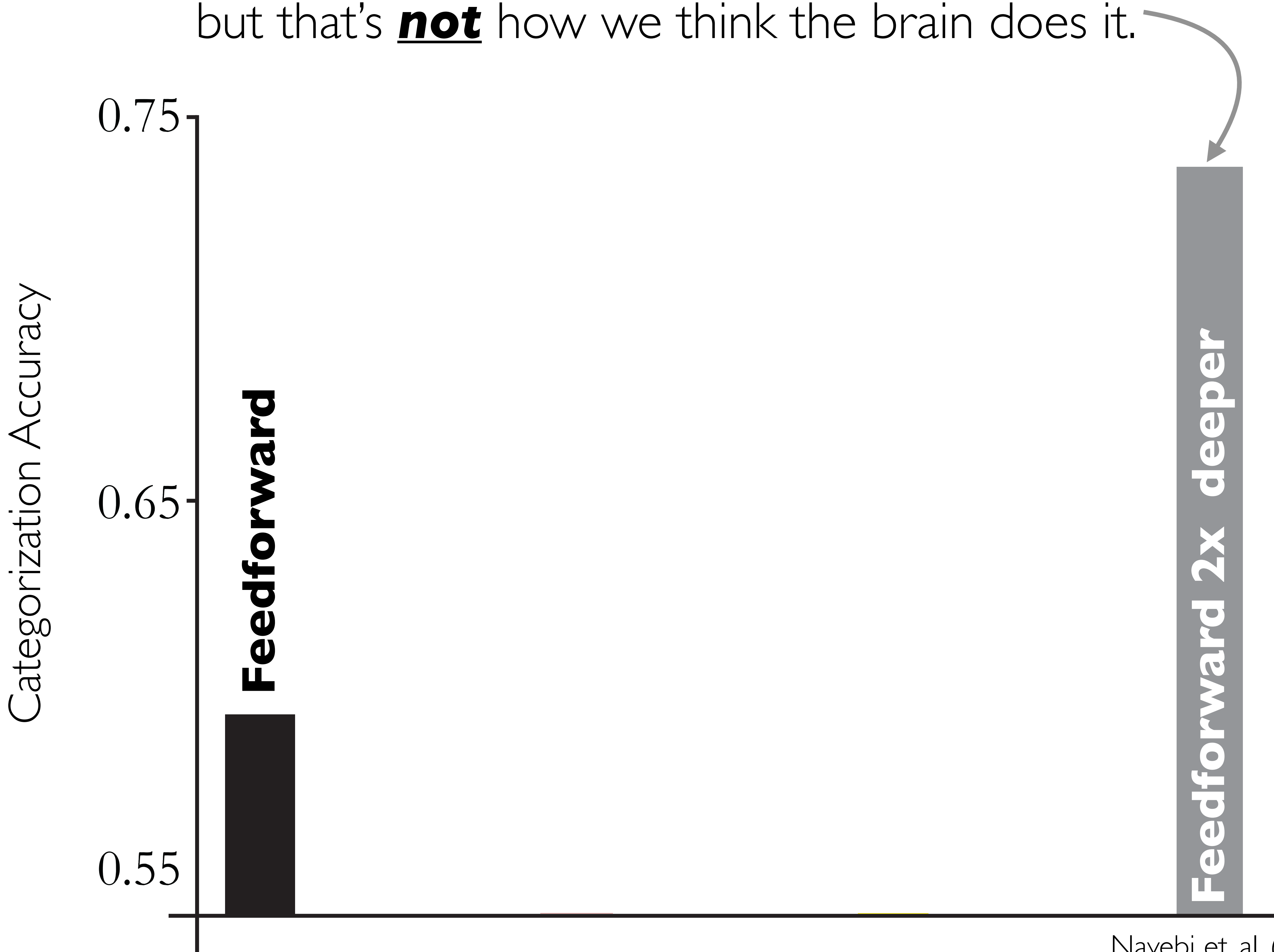
Improving ImageNet Performance with ConvRNNs



Nayebi et. al. (NeurIPS 2018)

Improving ImageNet Performance with ConvRNNs

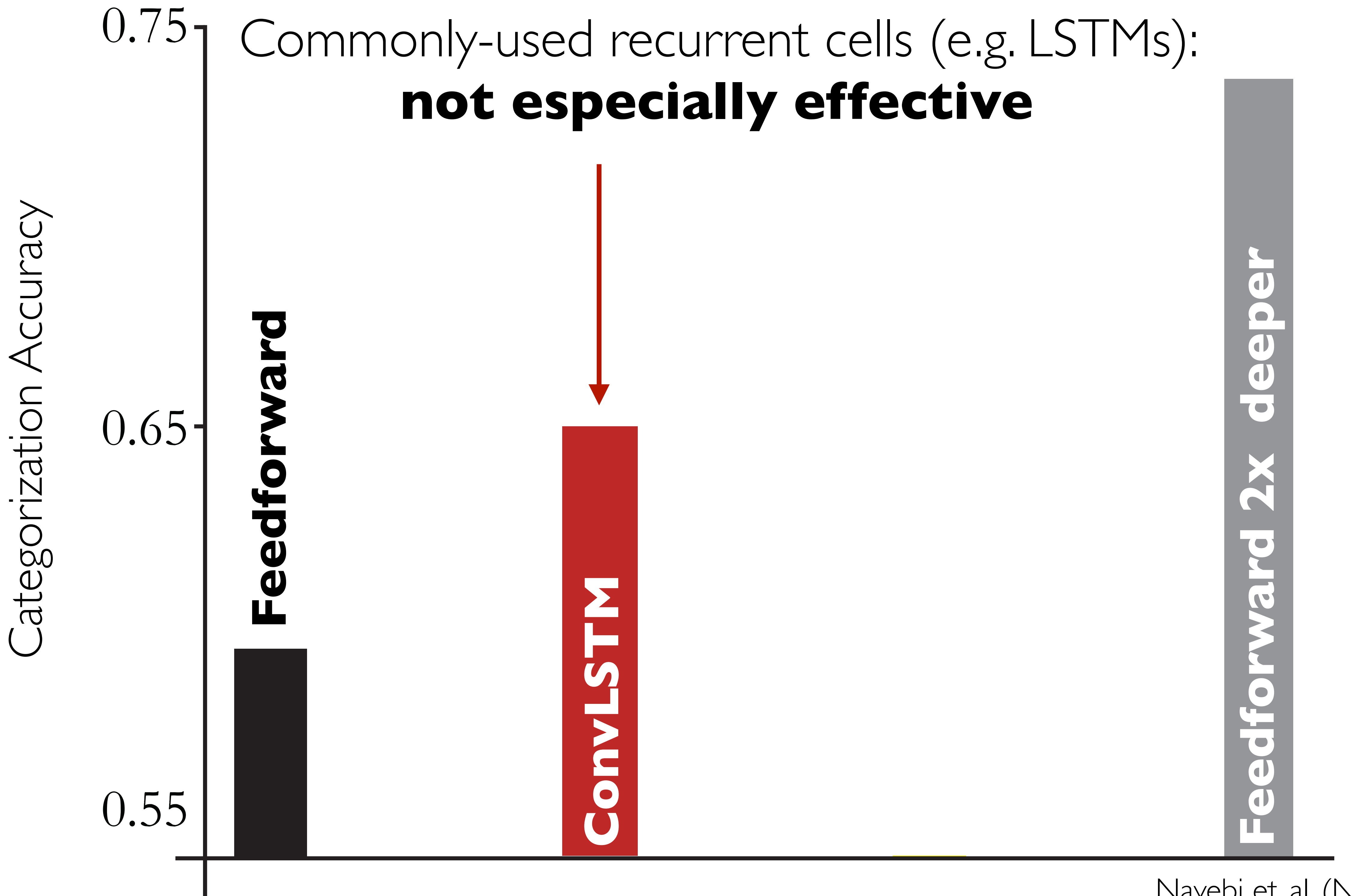
You can get better at performance with more layers / parameters, but that's ***not*** how we think the brain does it.



Nayebi et. al. (NeurIPS 2018)

Improving ImageNet Performance with ConvRNNs

You can get better at performance with more layers / parameters, but that's ***not*** how we think the brain does it.



Nayebi et. al. (NeurIPS 2018)

Improving ImageNet Performance with ConvRNNs

With better recurrent cells, substantial performance improvements on ImageNet

Improving ImageNet Performance with ConvRNNs

With better recurrent cells, substantial performance improvements on ImageNet

Two useful principles:

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

(2) bypassing = when recurrent cell is in 0 state, input is unchanged ("performance preserving")

Improving ImageNet Performance with ConvRNNs

With better recurrent cells, substantial performance improvements on ImageNet

Two useful principles:

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

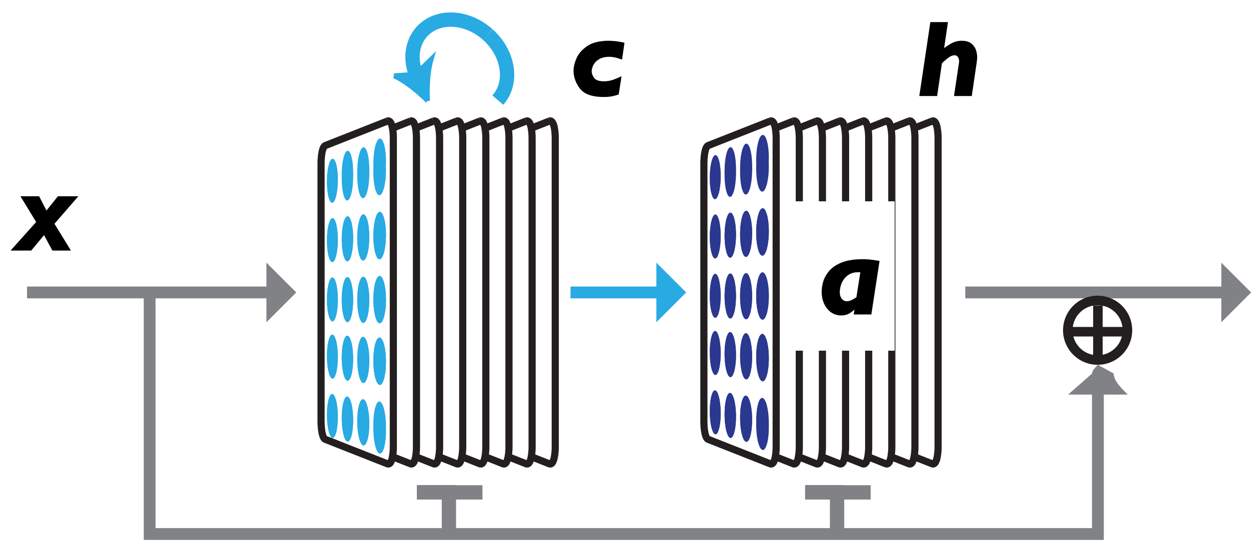
(2) bypassing = when recurrent cell is in 0 state, input is unchanged ("performance preserving")

SimpleRNN has **(2)** but not **(1)**

Standard Long Short-Term Memory (LSTM) has **(1)** but not **(2)**

Improving ImageNet Performance with ConvRNNs

"Resnet-Like" Unit

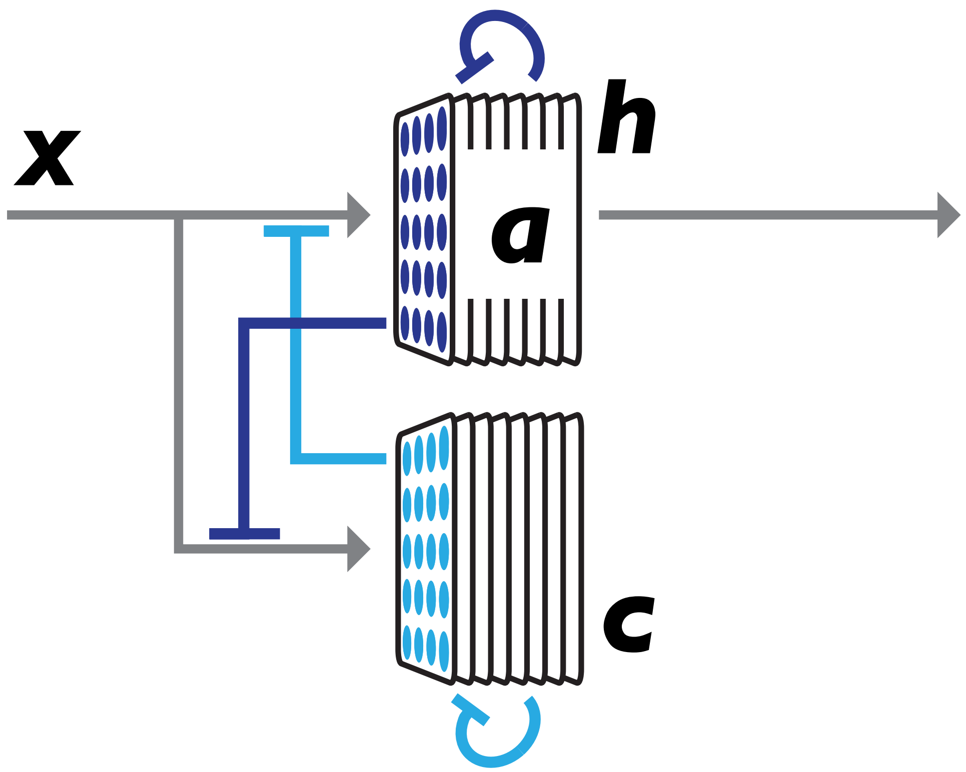


$$\mathbf{a}^{t+1} = \mathbf{x}^t + \sigma(W_{xh} \circledast \mathbf{x}^t)\mathbf{h}^t + W_{ch} \circledast \mathbf{c}^t$$
$$\mathbf{h}^t = f[\mathbf{a}^t]$$

$$\tilde{\mathbf{c}}^{t+1} = \sigma(W_{xc} \circledast \mathbf{x}^t) \cdot \mathbf{c}^t + W_{xc}\mathbf{x}^t + W_{cc} \circledast \mathbf{c}^T$$
$$\mathbf{c}^t = f[\tilde{\mathbf{c}}^t]$$

Improving ImageNet Performance with ConvRNNs

"Reciprocal Gated" Unit



$$\mathbf{a}^{t+1} = (1 - \sigma(W_{ch} \circledast \mathbf{c}^{t})) \cdot \mathbf{x}^{t}$$ "gated input"

$$+ (1 - \sigma(W_{hh} \circledast \mathbf{h}^{t})) \cdot \mathbf{h}^{t}$$ "gated memory"

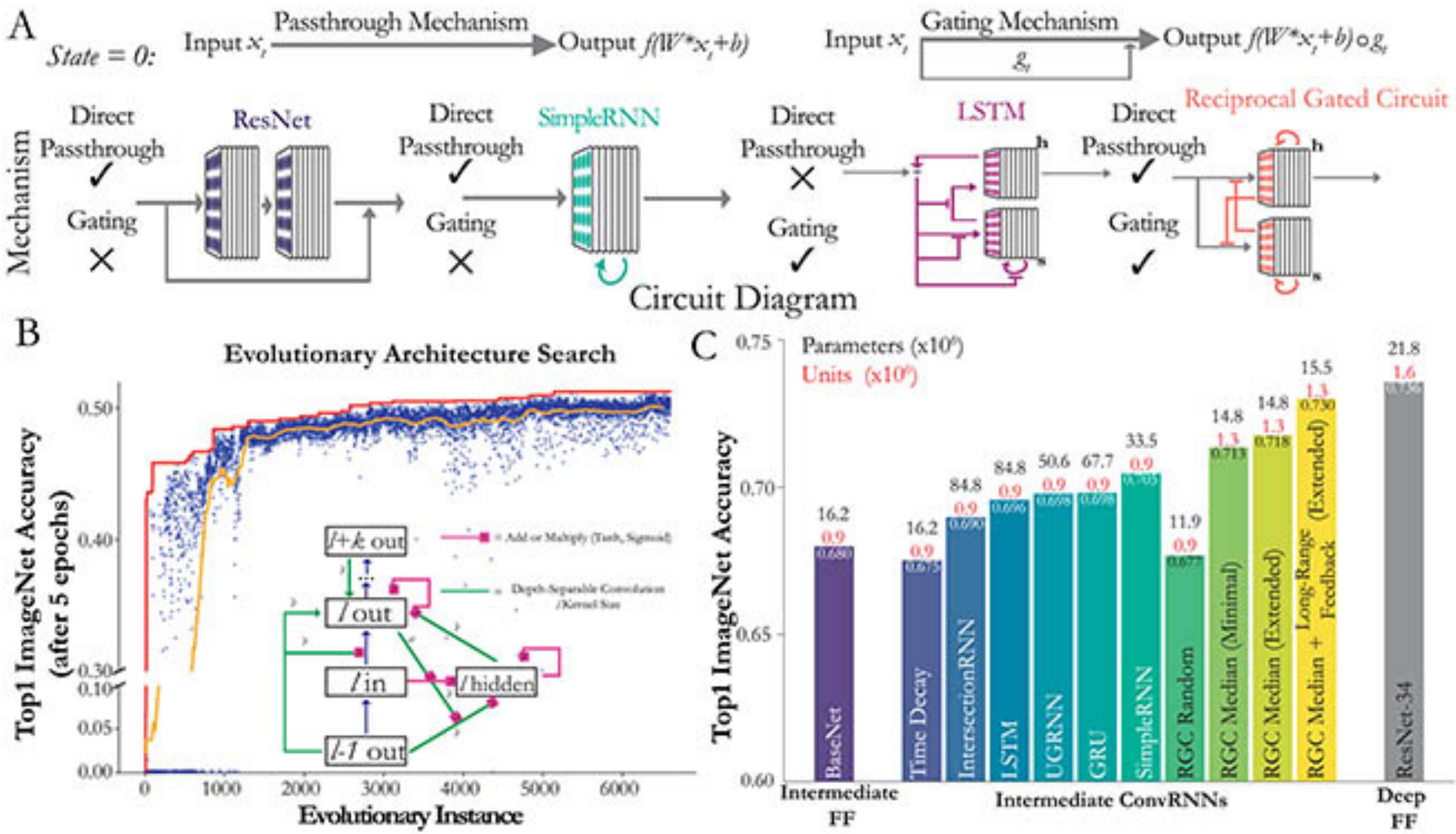
$$\mathbf{h}^{t} = f[\mathbf{a}^{t}]$$

$$\tilde{\mathbf{c}}^{t+1} = (1 - \sigma(W_{hc} \circledast \tilde{\mathbf{c}}^{t})) \cdot \mathbf{x}^{t}$$ reciprocal structure

$$+ (1 - \sigma(W_{cc} \circledast \mathbf{c}^{t})) \cdot \mathbf{c}^{t}$$

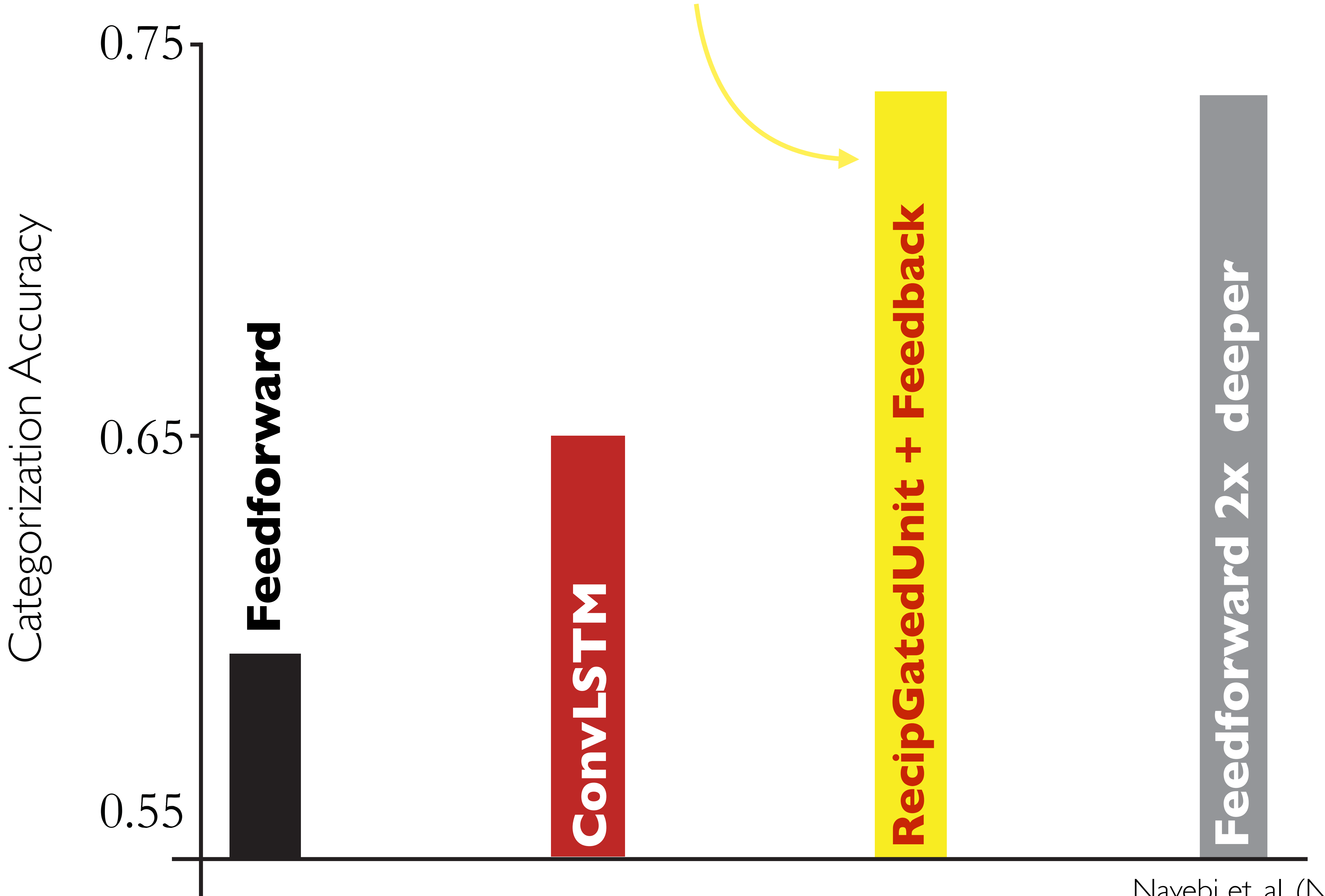
$$\mathbf{c}^{t} = f[\tilde{\mathbf{c}}^{t}]$$

Improving ImageNet Performance with ConvRNNs



Improving ImageNet Performance with ConvRNNs

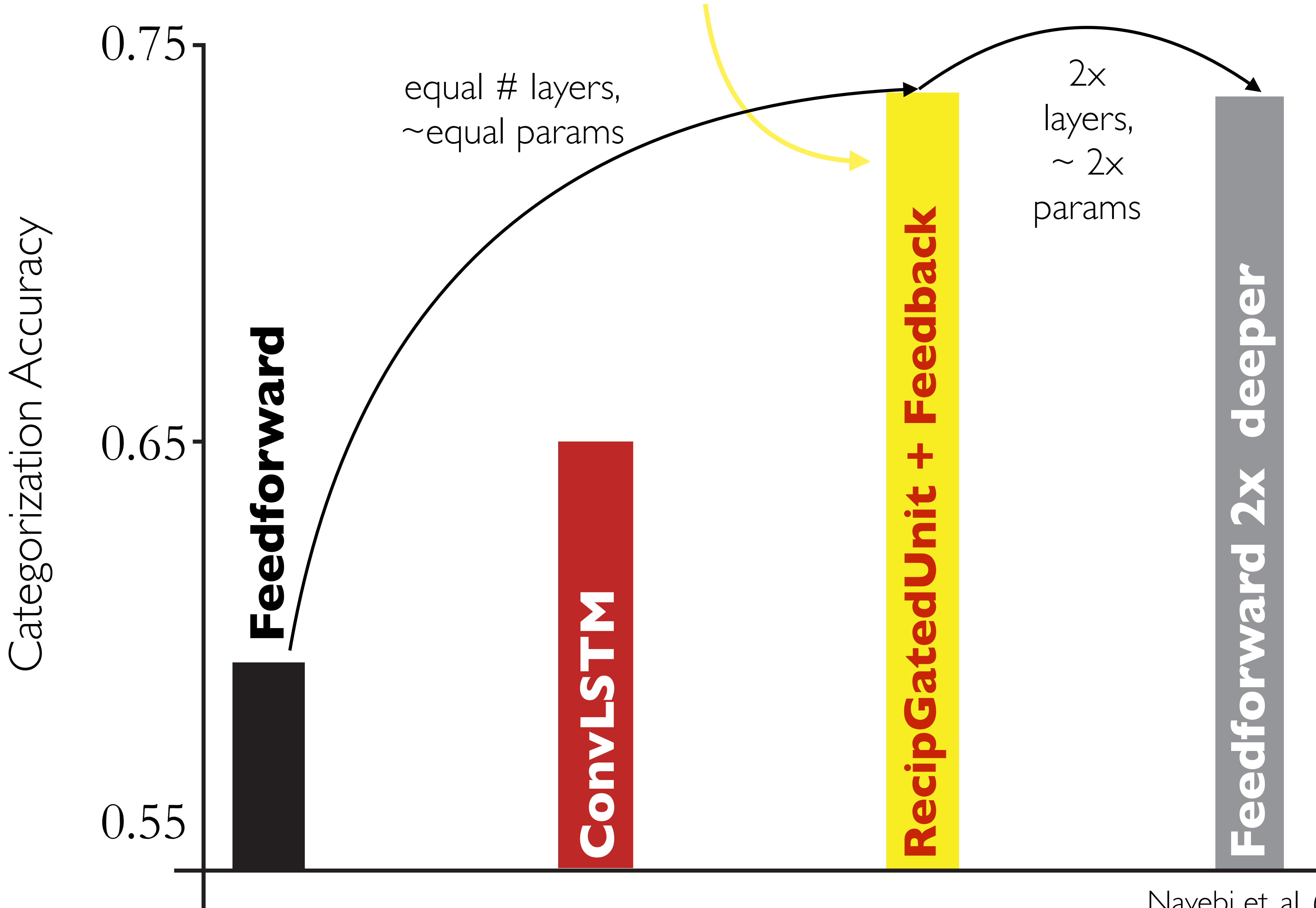
ConvRNNs, with correct local recurrence & long-range feedback
can effectively convert "space" into "time"



Nayebi et. al. (NeurIPS 2018)

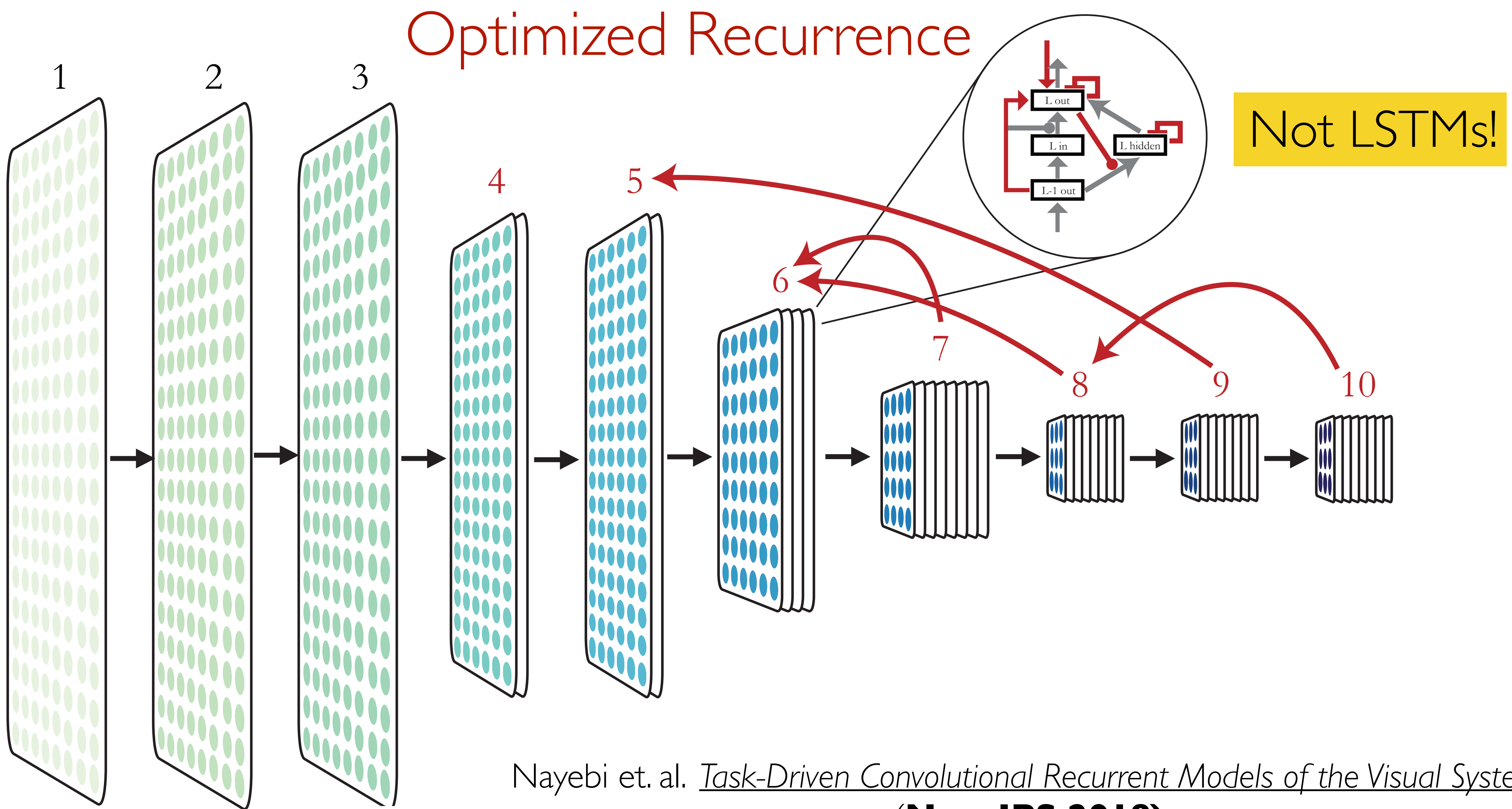
Improving ImageNet Performance with ConvRNNs

ConvRNNs, with correct local recurrence & long-range feedback
can effectively convert "space" into "time"



Nayebi et. al. (NeurIPS 2018)

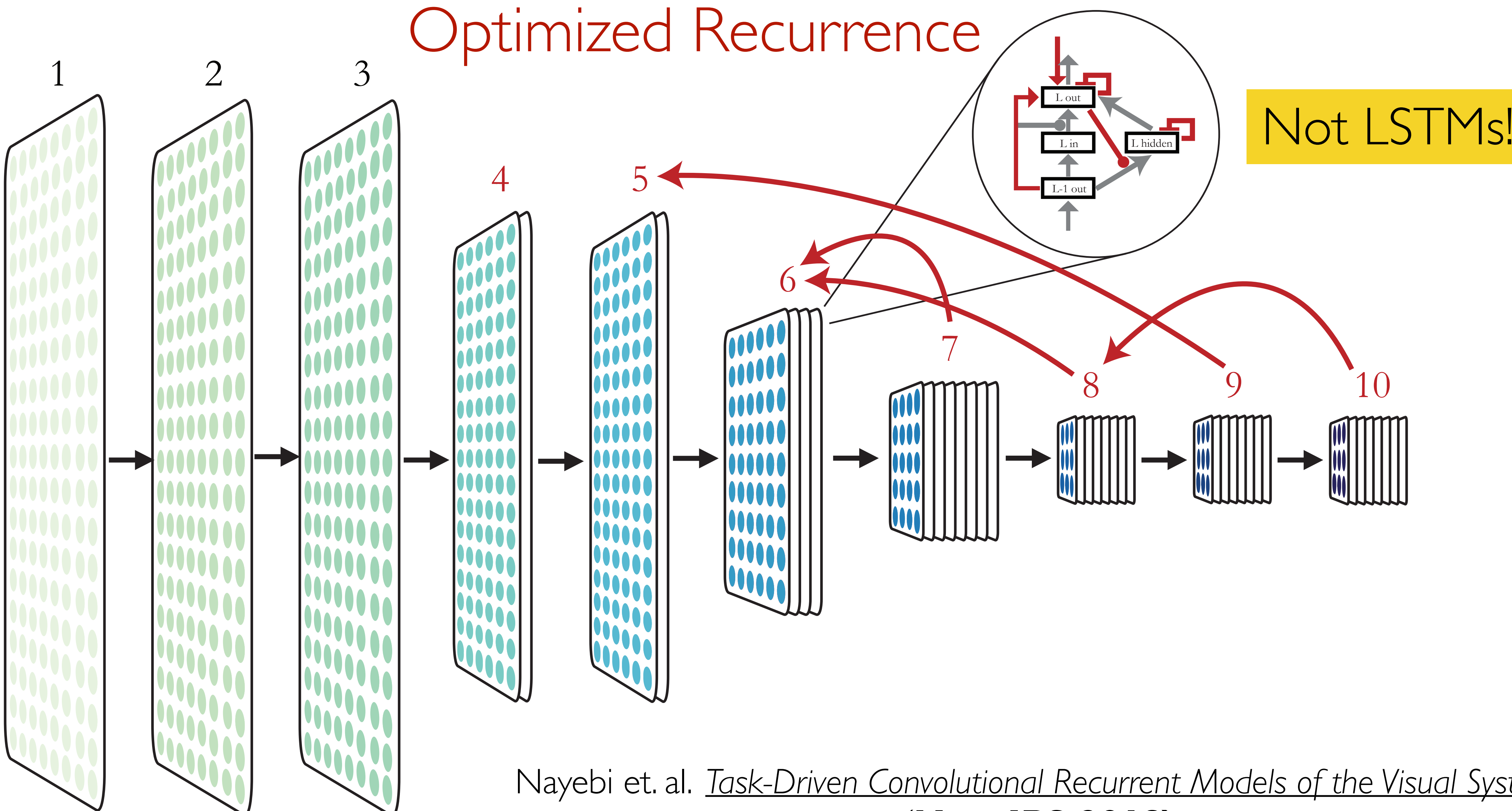
ConvRNNs as Models of Neural Dynamics



Nayebi et. al. *Task-Driven Convolutional Recurrent Models of the Visual System.* (**NeurIPS 2018)**

1) improved ImageNet performance

ConvRNNs as Models of Neural Dynamics

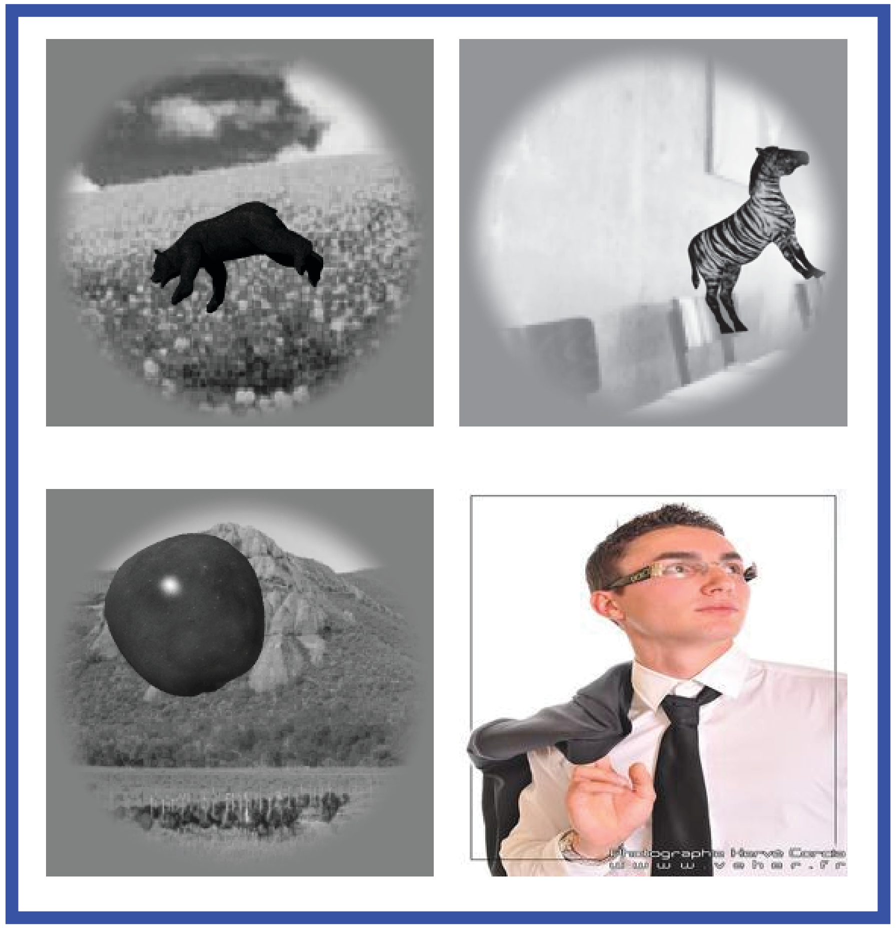


Nayebi et. al. *Task-Driven Convolutional Recurrent Models of the Visual System.* (**NeurIPS 2018**)

1) improved ImageNet performance

2) predictions of ***neural dynamics*** in visual system?

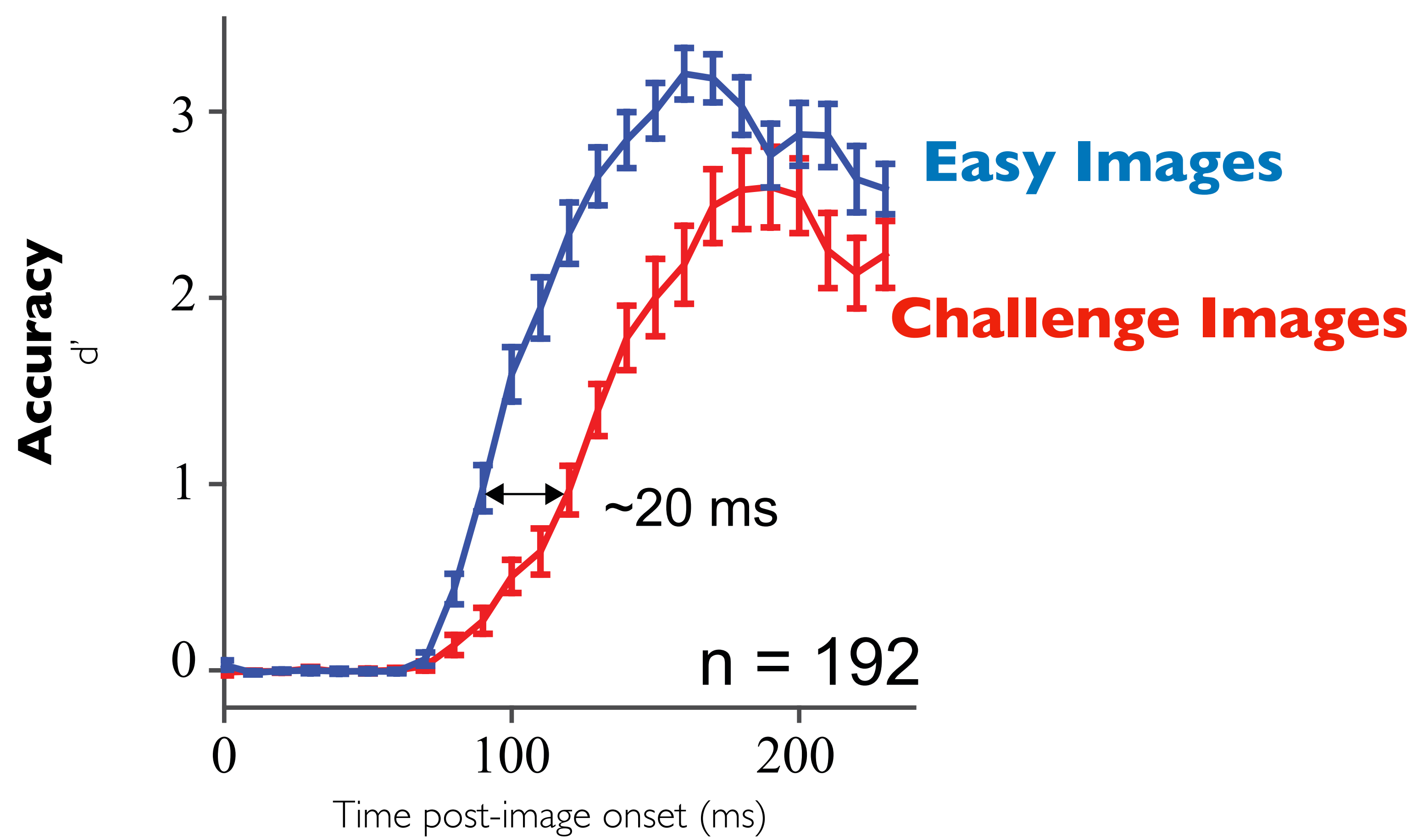
Challenge Images **Easy Images**



Kar et. al. (2017)

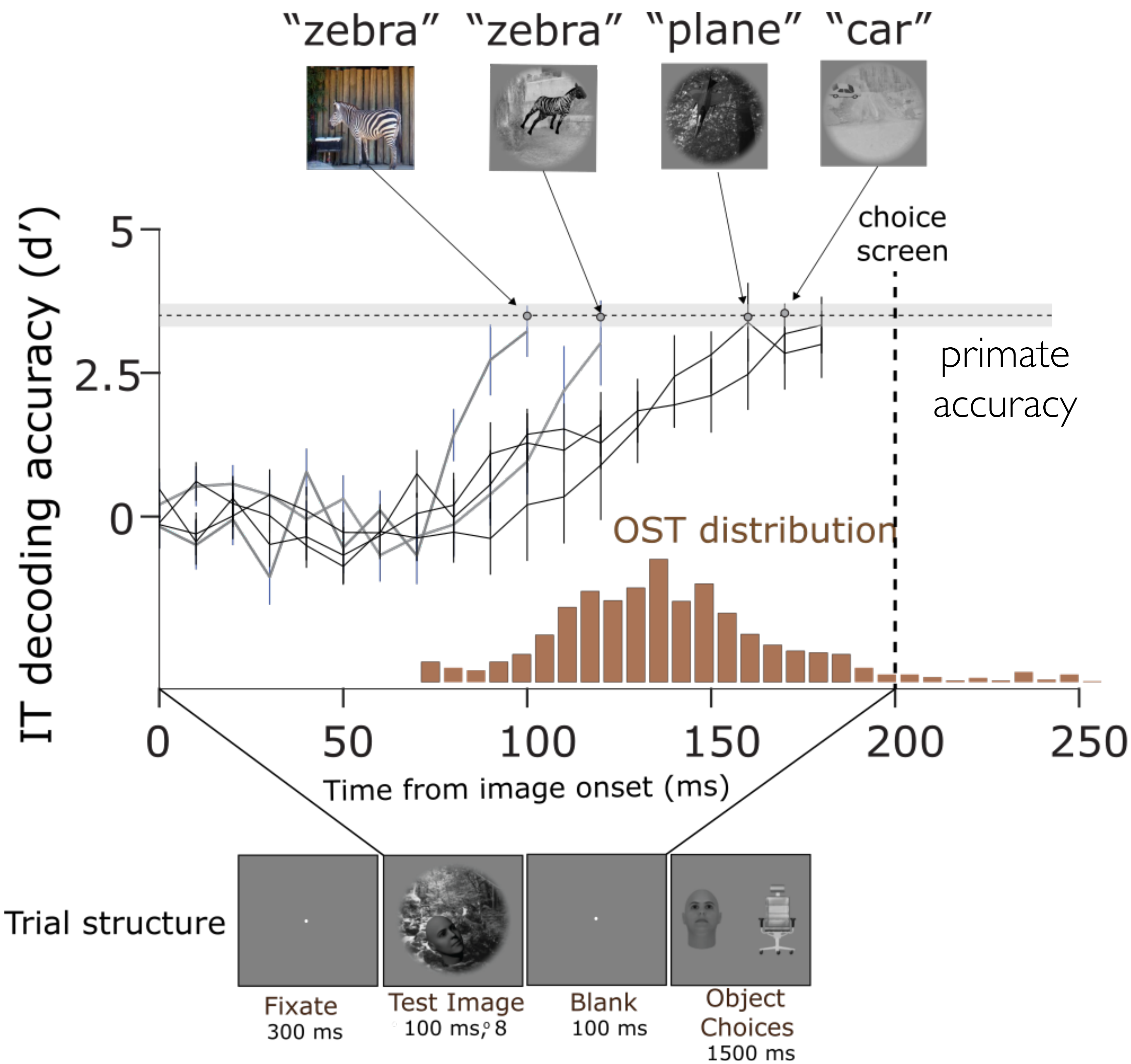
Challenge Images

Easy Images

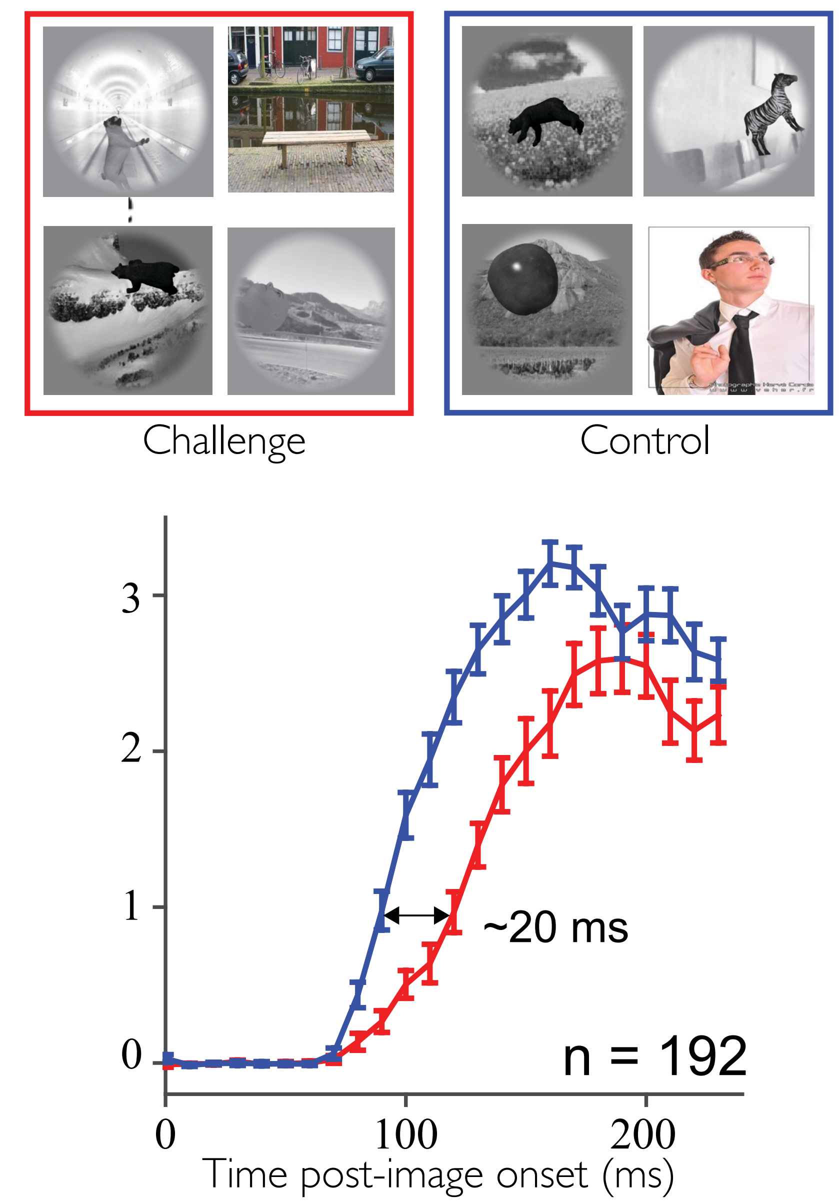


Kar et. al. (2017)

IT population dynamics reveal that each image is solved at a (slightly) different time

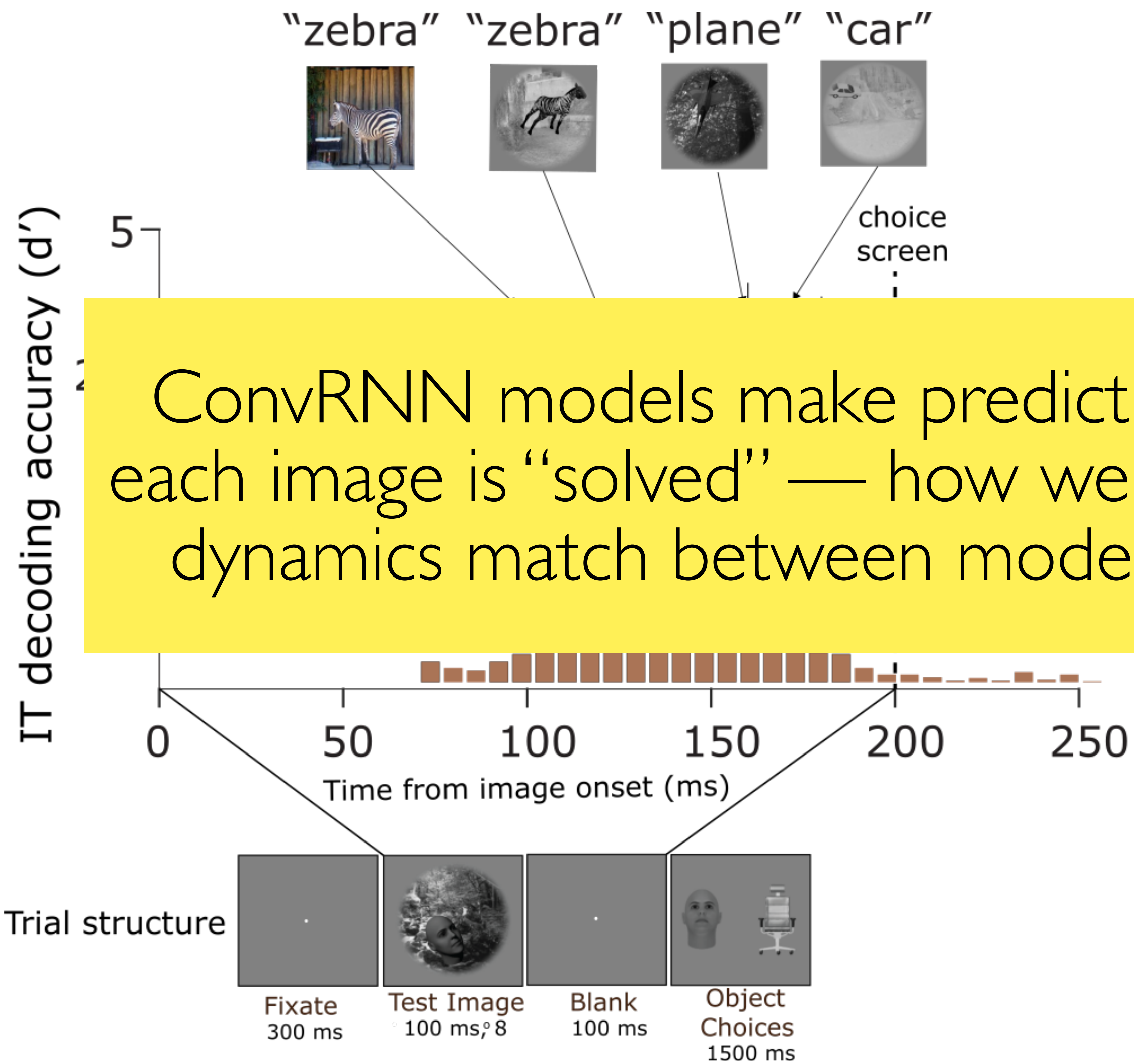


Hard images best decoded in late dynamics

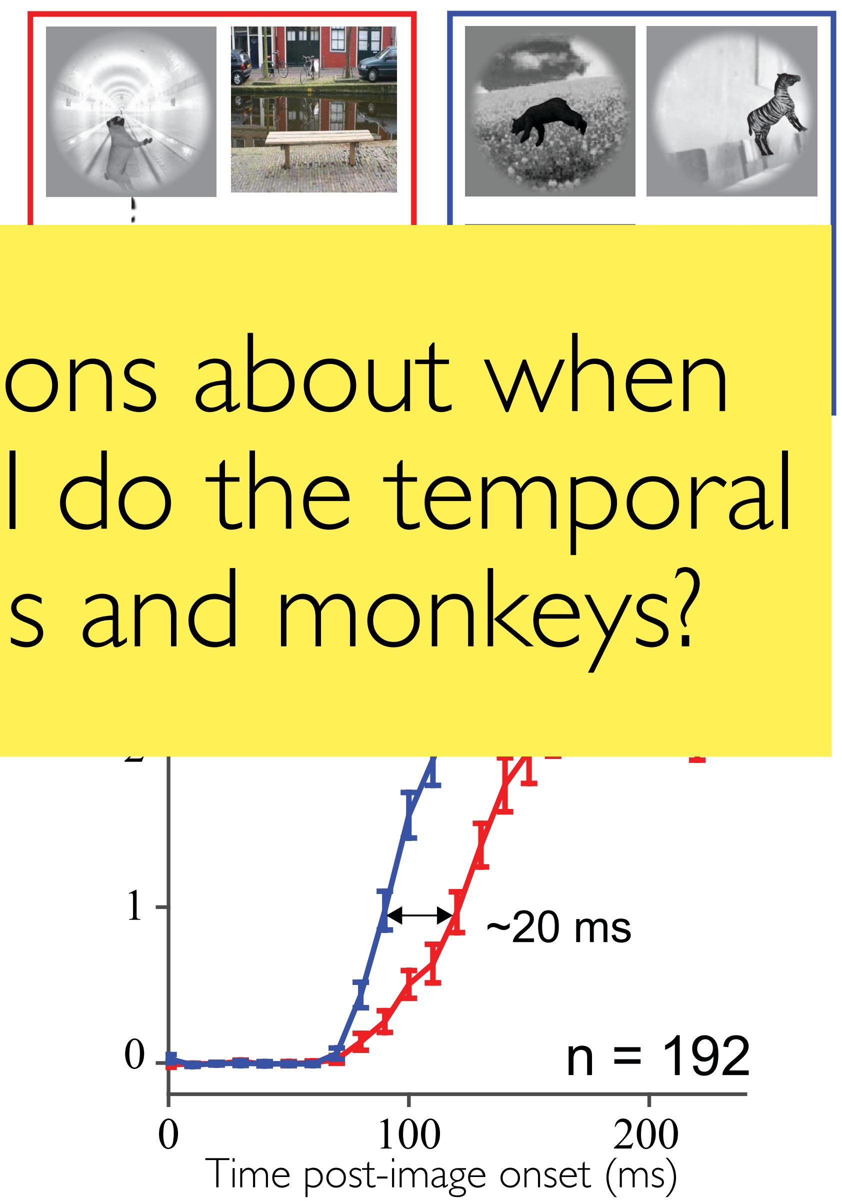


Kar et. al. (2017)

IT population dynamics reveal that each image is solved at a (slightly) different time



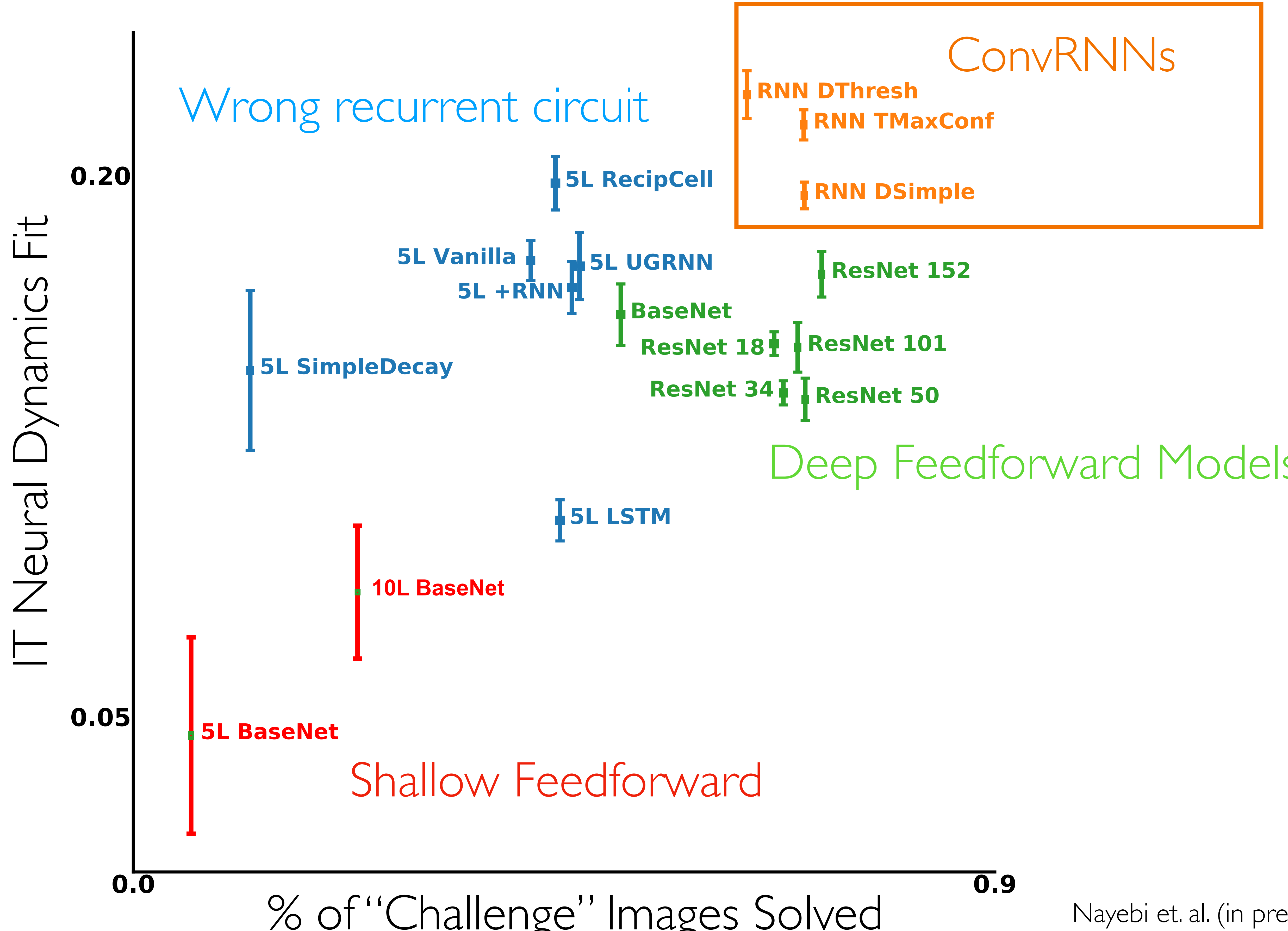
Hard images best decoded in late dynamics



ConvRNN models make predictions about when each image is "solved" — how well do the temporal dynamics match between models and monkeys?

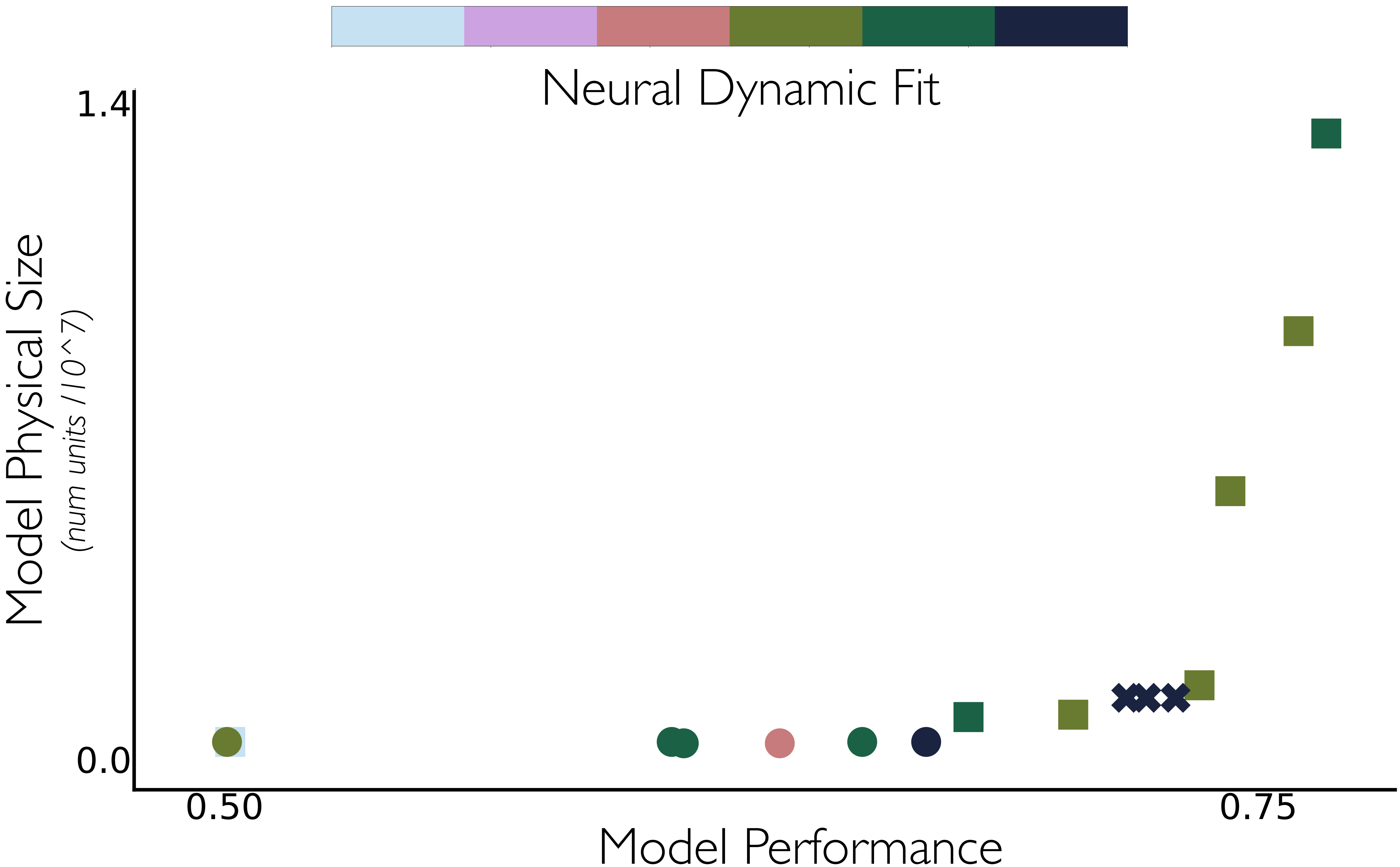
Kar et. al. (2017)

ConvRNNs as Models of Neural Dynamics



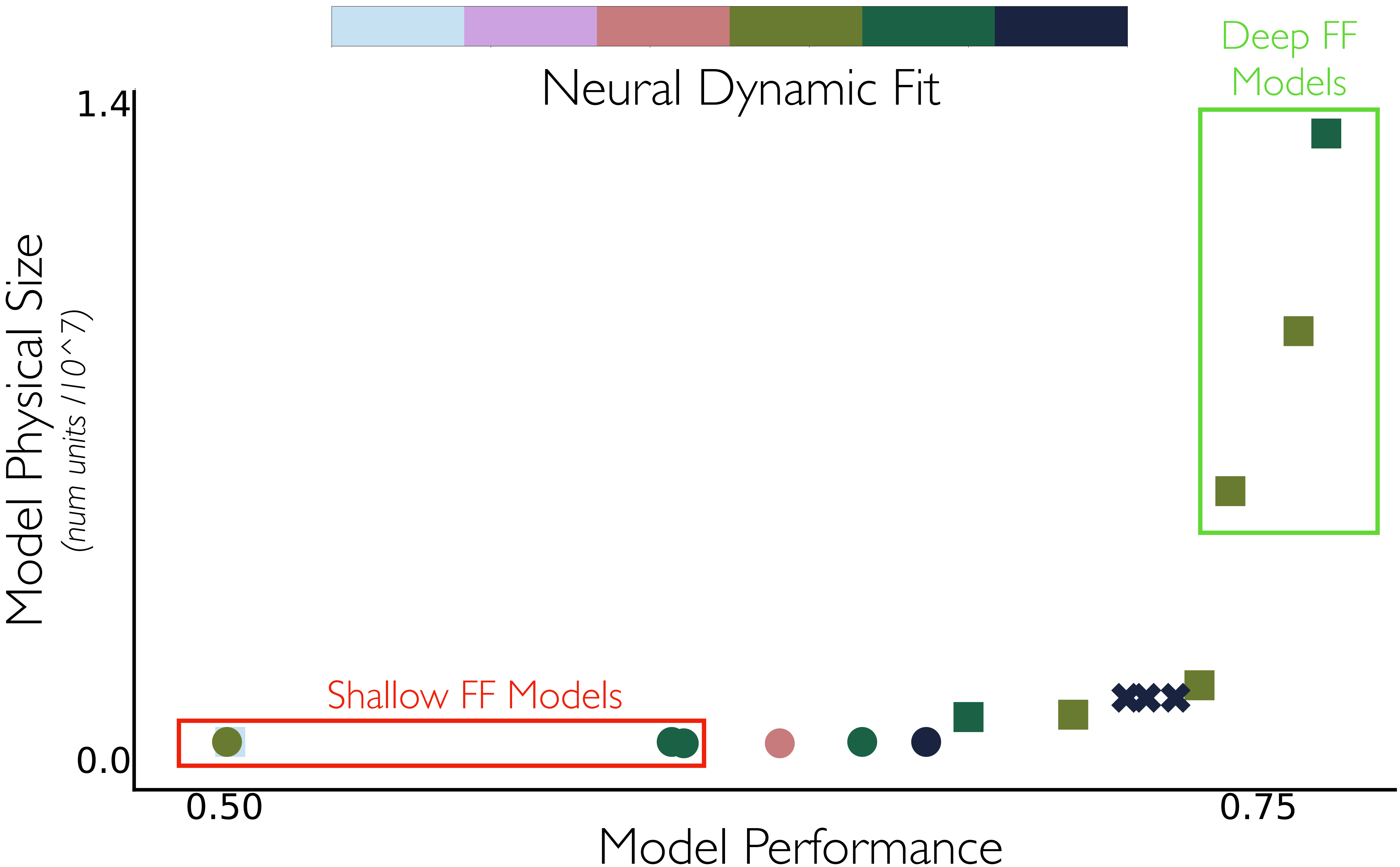
Nayebi et. al. (in prep)

ConvRNNs as Models of Neural Dynamics



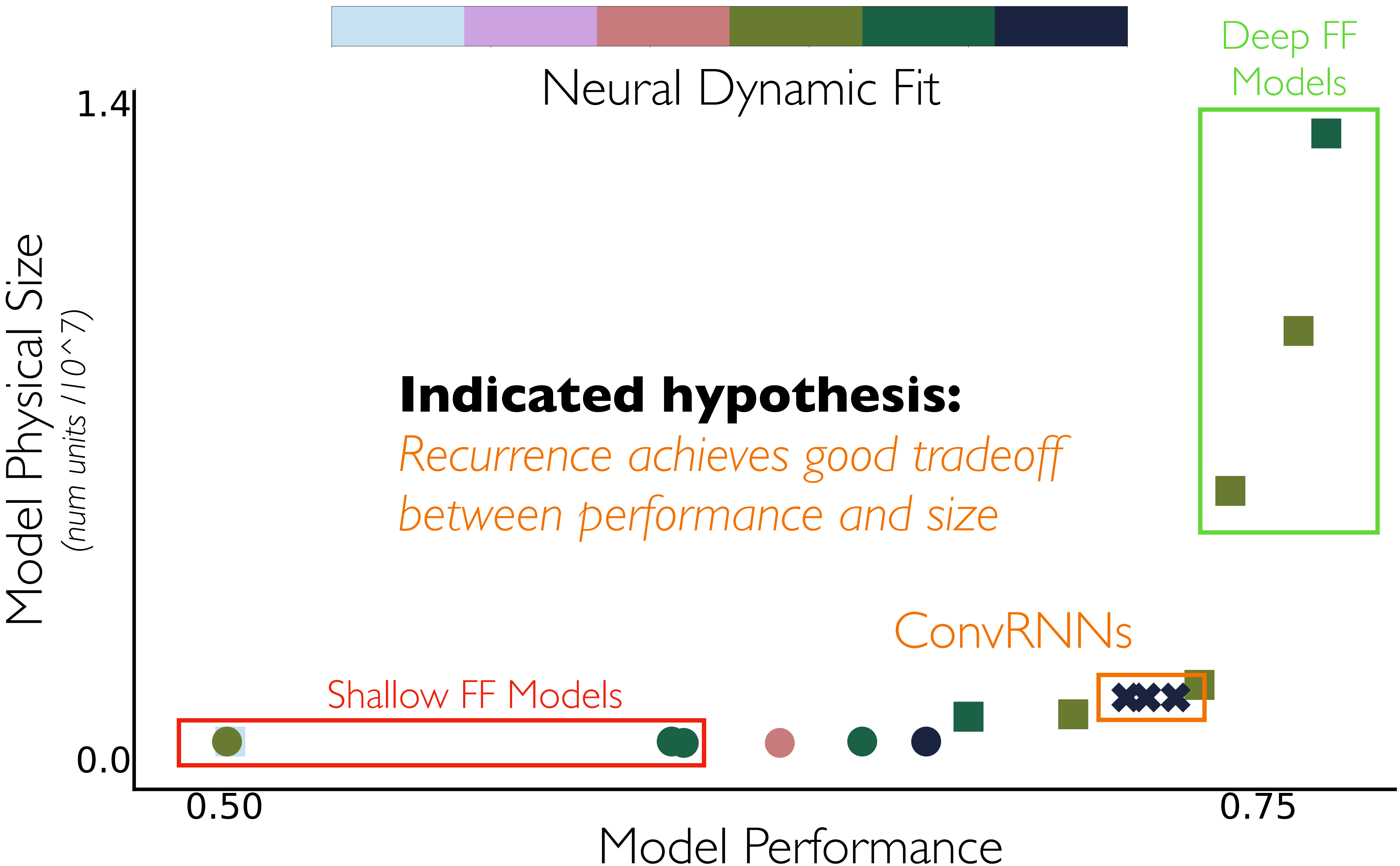
Nayebi et. al. (in prep)

ConvRNNs as Models of Neural Dynamics



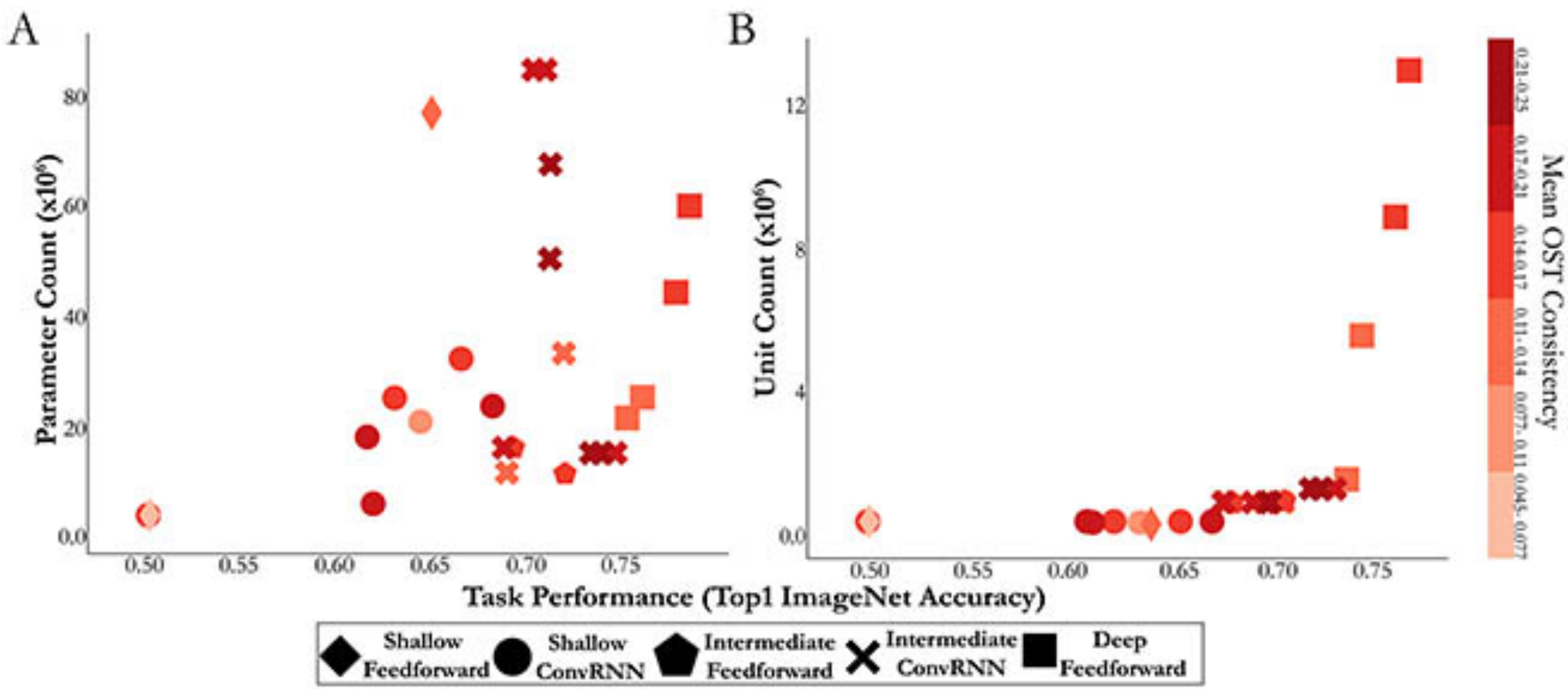
Nayebi et. al. (in prep)

ConvRNNs as Models of Neural Dynamics



Nayebi et. al. (in prep)

ConvRNNs as Models of Neural Dynamics



Models

Accepted as a workshop contribution at ICLR 2015

ATTENTION FOR FINE-GRAINED CATEGORIZATION

Pierre Sermanet, Andrea Frome, Esteban Real
Google, Inc.
{sermanet,afrome,ereal,}@google.com

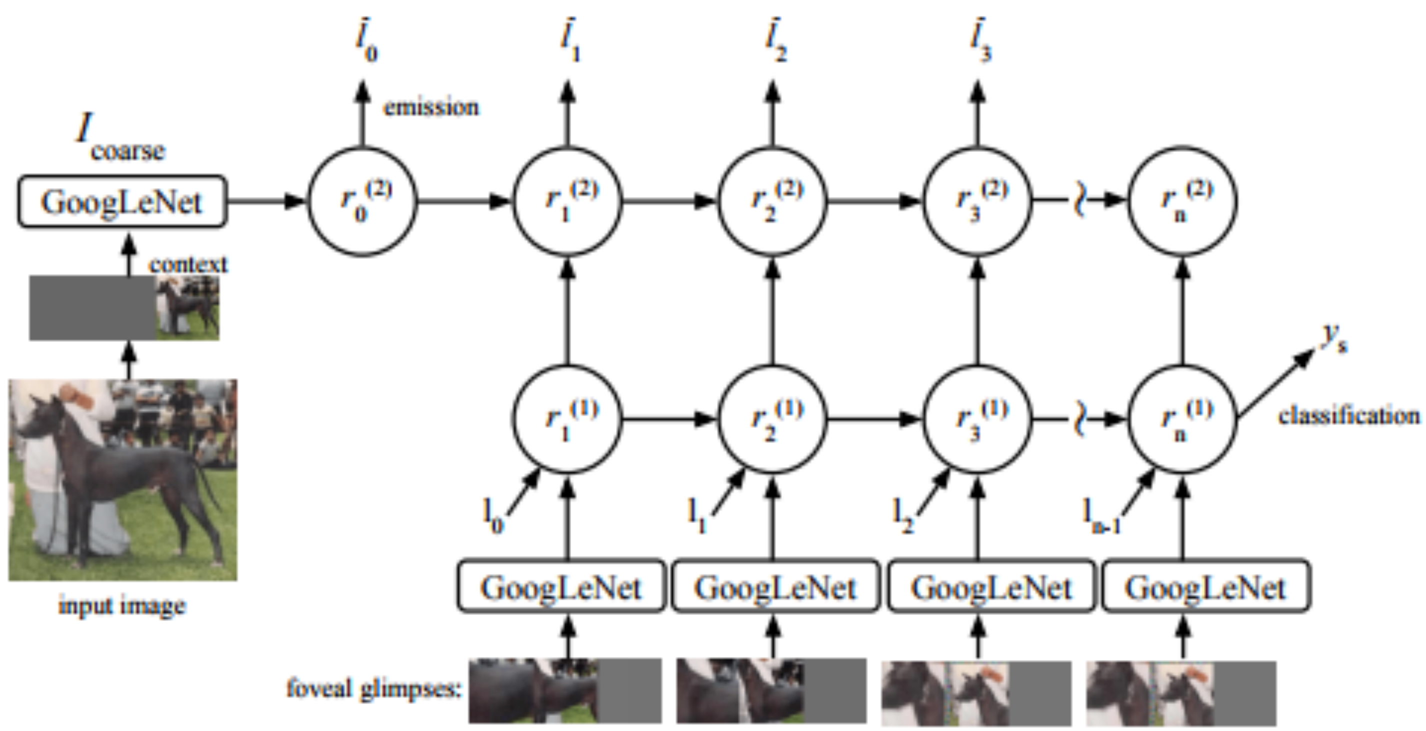


Figure 2: Diagram of the model. The grayed-out boxes denote resolutions not in use; in our experiments the context is always a low-resolution patch, while each glimpse can be any combination of the low-, medium-, and high-resolution patches.

Models

Accepted as a workshop contribution at ICLR 2015

ATTENTION FOR FINE-GRAINED CATEGORIZATION

Pierre Sermanet, Andrea Frome, Esteban Real
Google, Inc.
{sermanet,afrome,ereal,}@google.com

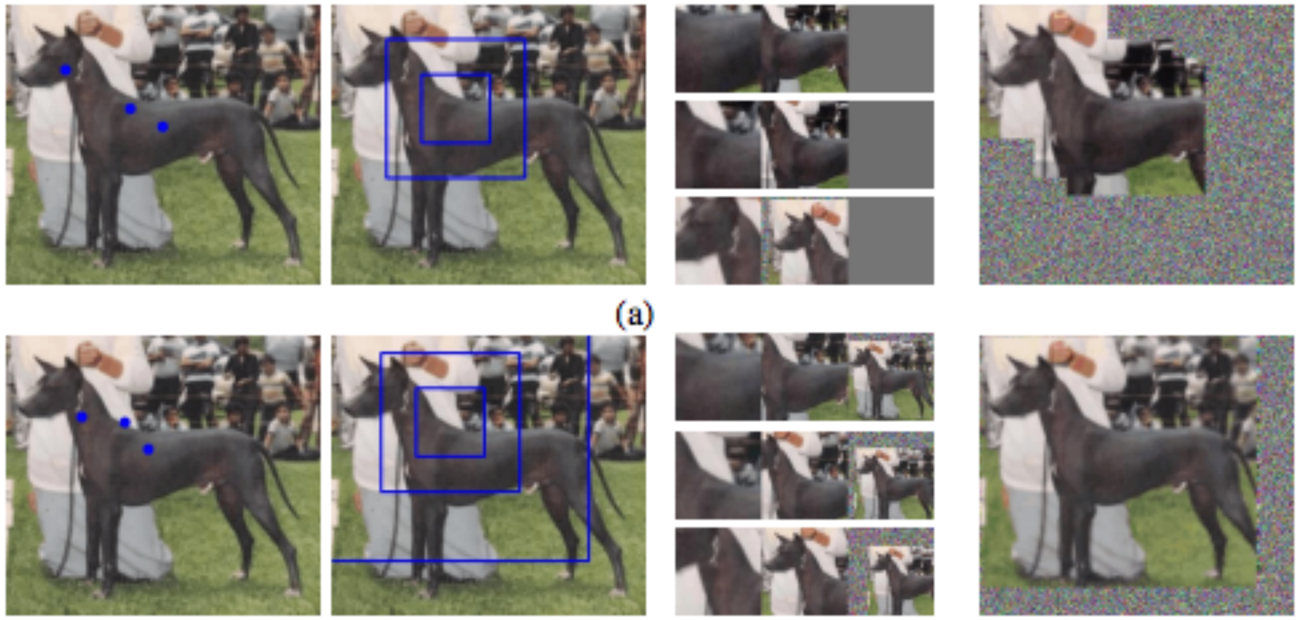
Table 1: Results on Stanford Dogs for (a) our RNN model and (b) our GoogLeNet baselines and previous state-of-the-art results, measured by mean accuracy percentage (mA) as described in Chai et al. (2013). The GoogLeNet baseline models were pre-trained on the de-duped ILSVRC 2012 training set and fine-tuned with the Stanford Dogs training set. Results marked with a star indicate use of tight ground truth bounding boxes around the dogs in training and testing.

# glimpses	1	2	3
high res only	43.5	48.3	49.6
medium res only	70.1	72.3	72.8
low res only	70.3	70.1	70.7
high+medium res	70.7	72.6	72.7
3-resolution	76.3	76.5	**76.8**

(a)

Yang et al. (2012)*	38.0
Chai et al. (2013)*	45.6
Gavves et al. (2013)*	50.1
GoogLeNet 96×96	58.8
GoogLeNet 224×224	**75.5**

(b)



(a)

(b)

Figure 3: Visualizations of 2-resolution (a) and 3-resolution (b) glimpses on an image from our validation set, with learned fixation points. For each the glimpse images are in order, from top to bottom, and the box diagram corresponds to the second glimpse. The composite image is created from all three glimpses. The context image is not shown but is always the same resolution and size as the low-resolution glimpse patches shown in (b).

Task-Driven Models?

1.

A = *architecture class*

CNNs -> RNNs

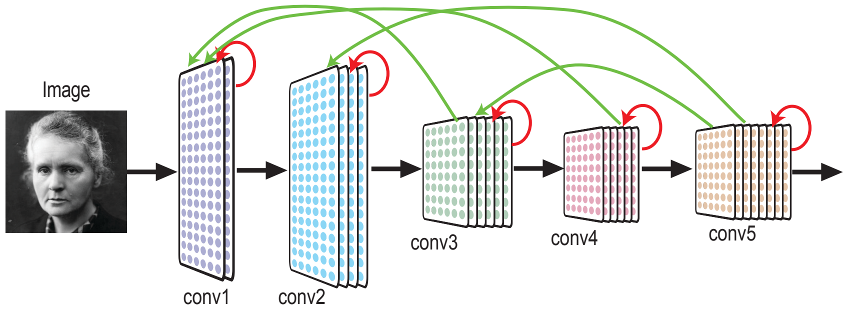
“*task*”

2.

L = *loss function* ***D*** = *dataset*

e.g. **Object Categorization**

What task(s) explain recurrences??



Hard Images
(e.g. *heavy occlusion*)

Time-accuracy tradeoff
(*be correct but fast*)

Temporal Goal
(e.g. *motion-based*)

Task-Driven Models?

1.

A = *architecture class*

CNNs -> RNNs

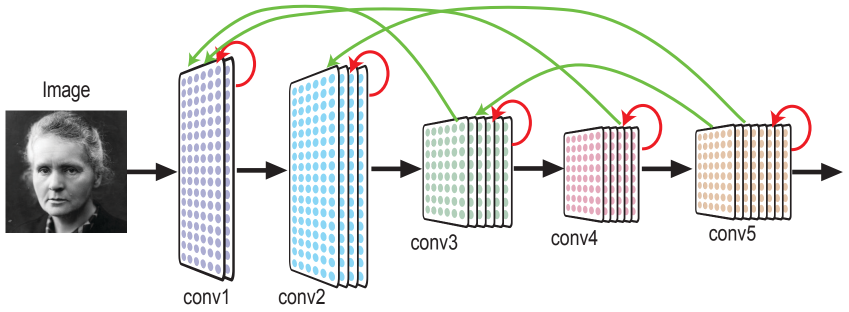
"task"

2.

L = *loss function* ***D*** = *dataset*

e.g. **Object Categorization**

What task(s) explain recurrences??



Hard Images
(e.g. heavy occlusion)

Time-accuracy tradeoff
(be correct but fast)

Temporal Goal
(e.g. motion-based)

very different possibility: actually, recurrence not used on-line,
instead: "just" implementing learning

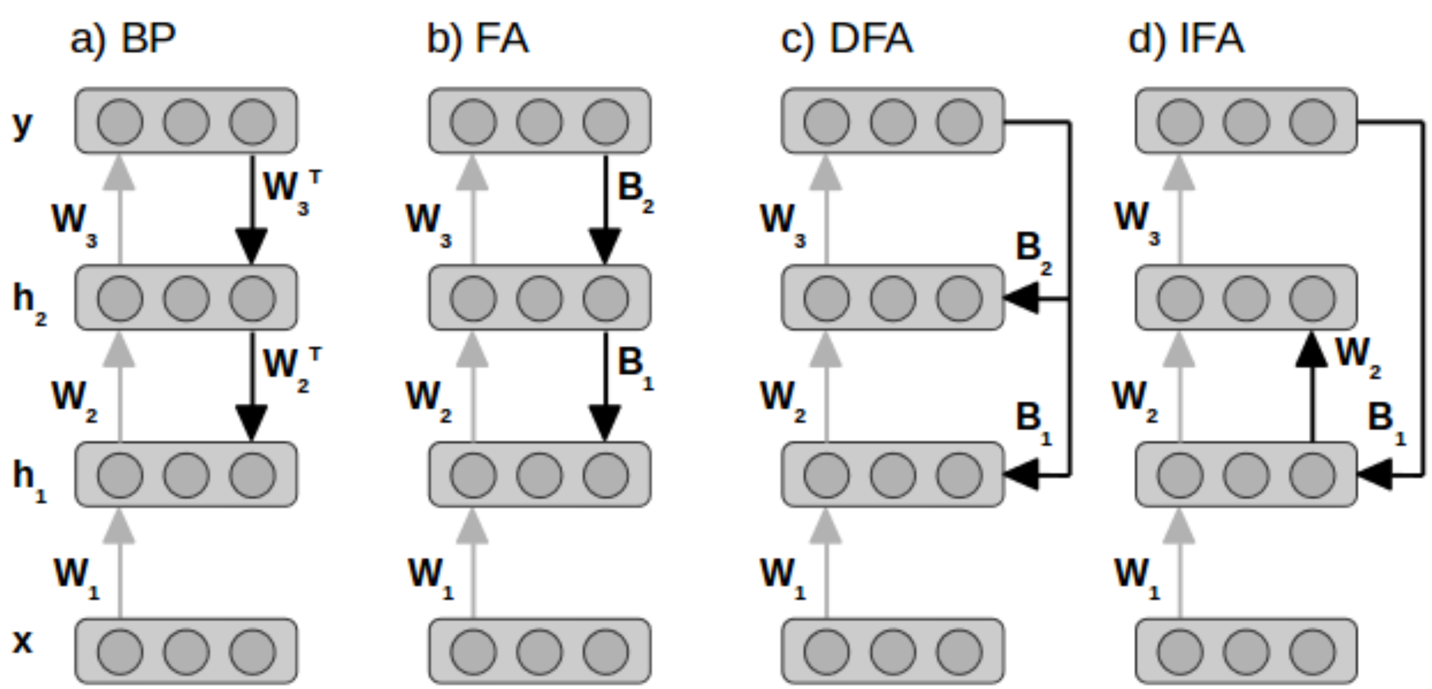
Biological learning

Implementing Learning

Direct Feedback Alignment Provides Learning in Deep Neural Networks

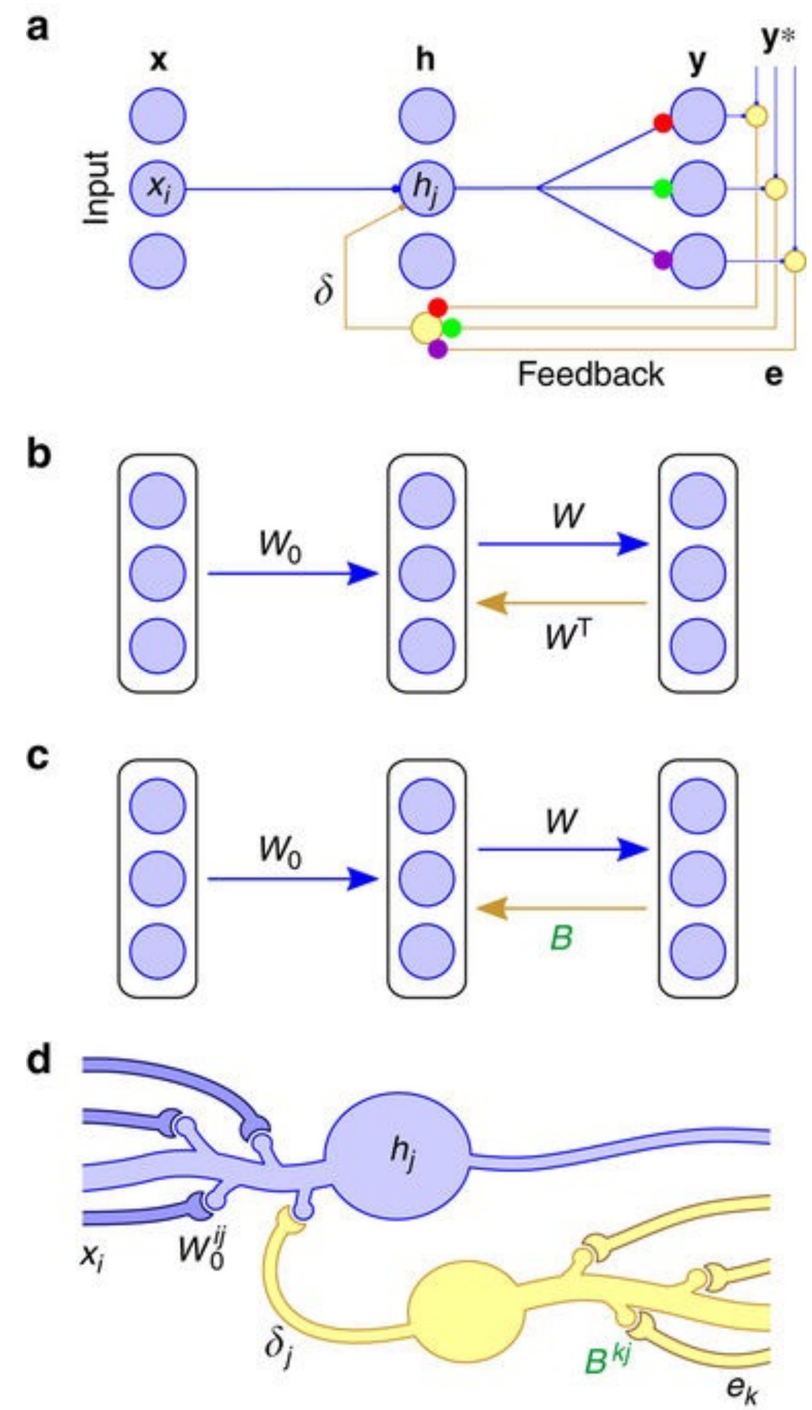
Arild Nøkland

(Submitted on 6 Sep 2016 (v1), last revised 21 Dec 2016 (this version, v5))



Random synaptic feedback weights support error backpropagation for deep learning

Timothy P. Lillicrap, Daniel Cownden, Douglas B. Tweed & Colin J. Akerman



Big Problems in Each Area

***bad** = obviously deeply wrong as model of the brain or behavior

1. ✗ ***bad***

***A** = architecture class*

ConvRNNs

2.

***T** = task/objective*

e.g. **Object Categorization**

3.

***D** = dataset*

e.g. **ImageNet**

4.

***L** = learning rule*

e.g. **Arch. Srch. + Grad. Desc.**

SOLUTION

RECURRENCE and FEEDBACK

Big Problems in Each Area

*✓ok = we've really nailed it

*bad = obviously deeply wrong

1. ~~*✓ok~~

A = *architecture class*

ConvRNNs

2.

T = *task/objective*

e.g. **Object Categorization**

3.

D = *dataset*

e.g. **ImageNet**

4.

L = *learning rule*

e.g. **Arch. Srch. + Grad. Desc.**

SOLUTION

RECURRENCE and FEEDBACK

Big Problems in Each Area

*✓**ok** = we've really nailed it
*✓**ok-ish** = **harder to reject out of hand**
***bad** = obviously deeply wrong

1. *✓**ok-ish**

***A** = architecture class*

ConvRNNs

2.

***T** = task/objective*

e.g. **Object Categorization**

3.

***D** = dataset*

e.g. **ImageNet**

4.

***L** = learning rule*

e.g. **Arch. Srch. + Grad. Desc.**

SOLUTION

RECURRENCE and FEEDBACK

Big Problems in Each Area

*✓**ok** = we've really nailed it
*✓**ok-ish** = **harder to reject out of hand**
***bad** = obviously deeply wrong

1. ***✓ok-ish**

***A** = architecture class*

ConvRNNs

2. **✓ok**

***T** = task/objective*

e.g. **Object Categorization**

3. ***✓ok-ish**

***D** = dataset*

e.g. **ImageNet**

4. **✗bad**

***L** = learning rule*

e.g. **Arch. Srch.** + **Grad. Desc.**

SOLUTION

RECURRENCE and FEEDBACK

SELF-SUPERVISION WORKS GREAT!

*CAN HANDLE REAL VIDEOSTREAMS TO *SOME* EXTENT*